



HHS Public Access

Author manuscript

New Dir Eval. Author manuscript; available in PMC 2018 June 25.

Published in final edited form as:

New Dir Eval. 2018 ; 2018(158): 49–72. doi:10.1002/ev.20322.

Evaluative Thinking in Practice: The National Asthma Control Program

Leslie A. Fierro, Heather Codd, Sarah Gill, Phung K. Pham, Piper T. Grandjean Targos, and Maureen Wilce

Abstract

Although evaluative thinking lies at the heart of what we do as evaluators and what we hope to promote in others through our efforts to build evaluation capacity, researchers have given limited attention to measuring this concept. We undertook a research study to better understand how instances of evaluative thinking may present in practice-based settings—specifically within four state asthma control programs funded by the Centers for Disease Control and Prevention’s National Asthma Control Program. Through content analyses of documents as well as interviews and a subsequent focus group with four state asthma control programs’ evaluators and program managers we identified and defined twenty-two indicators of evaluative thinking. Findings provide insights about what practitioners may wish to look for when they intend to build evaluative thinking and the types of data sources that may be more or less helpful in such efforts.

As Patton describes in the first chapter of this volume, evaluative thinking has a long history within the field of evaluation. This volume demonstrates that evaluative thinking remains an important concept that continues to evolve. Despite its central importance to the field of evaluation, to our knowledge, few attempts have been made to operationalize and measure evaluative thinking in practice; excepting Cornell’s Office for Research on Evaluation (n.d.) set of data collection instruments on evaluative thinking and the *Evaluative Thinking Assessment Tool* published by the Bruner Foundation (Bruner Foundation, 2010). In addition, Patton has worked to clarify the “core elements of evaluative thinking” (Patton, this issue) but not with the aim of operationalization.

Operationalizing this construct is important for public health programs that devote significant resources to building evaluation capacity and wish to evaluate their efforts. The importance of operationalization may increase as public health programs turn purposeful attention to building evaluative thinking among their partners—a path of action some scholars have recently proposed (Buckley, Archibald, Hargraves, & Trochim, 2015). This is particularly true given that evaluative thinking presumably lies at the heart of what we do as evaluators and is certainly at least a portion of what we hope to promote in others through our efforts to build evaluation capacity (King, 2007). Even if specific quantitative measurement of this somewhat ambiguous construct is far off or infeasible at the moment, at a minimum, it seems important for evaluators to have some guidance about how to detect when evaluative thinking is present and how (or if) it is changing over time. In this chapter, we describe our efforts to better understand how instances of evaluative thinking may present in practice-based settings.

In performing this research, we operated from several underlying assumptions. Our first assumption, and perhaps most obvious, is that evaluative thinking can be operationalized. We assumed that “instances” of evaluative thinking would surface in existing grant-reporting documentation as well as in conversations with grantees. Evaluative thinking would not be captured by a single indicator, rather there would be a combination of indicators suggestive of evaluative thinking. We did not have assumptions about what this combination of indicators would look like, nor the relationships between them (if any), and allowed both to emerge organically through our efforts.

Second, we assumed that detection of instances of evaluative thinking would be aided greatly by including individuals with different backgrounds and experience on our team. To this end, we created a team comprised of six individuals who would naturally approach the detection of evaluative thinking from different angles—some from a practice-based lens and some from the lens of existing theory and scholarship in the area of evaluative thinking and related concepts such as evaluation capacity building (ECB). On the practice-based side, two of our team members have, combined, two decades of experience in providing evaluation technical assistance to the types of public health programs participating in the study. On the theory side, three team members are doctoral level graduate students working on research specific to evaluative thinking or ECB. One team member crossed these boundaries, with several years of experience developing and fostering the growth of evaluation capacity in the public health program of interest but currently working as a professor of evaluation performing research in evaluation capacity.

Finally, informed by the extensive literature on ECB (Labin, Duffy, Meyers, Wandersman, & Lesene, 2012), we operated from an underlying assumption that evaluative thinking is one of the many outcomes likely to arise from deliberate efforts to build evaluation capacity within organizations (Fierro, 2012). Thus, we assumed that where evaluation capacity exists evaluative thinking may also exist. As such, we explored the extent to which instances of evaluative thinking emerged within a national public health program with a long and documented history of commitment to ECB—the Centers for Disease Control and Prevention’s (CDC) National Asthma Control Program (NACP).

Evaluation Capacity Building in CDC’s National Asthma Control Program

A primary role for the NACP is funding state health departments throughout the United States to establish and maintain state-wide asthma programs that employ a range of interventions to improve asthma control and the quality of life of individuals who have asthma. Through this funding, awardees: (1) develop and maintain state-specific asthma surveillance systems that provide crucial data for understanding potential areas for targeting interventions, (2) engage partners in coordinated efforts to address state asthma goals and objectives, and (3) design and implement interventions to effectively manage and control asthma among people who live with asthma. Recognizing evaluation as an essential function of public health (State, Tribal, Local, & Territorial Public Health Professionals Gateway, n.d.) as well as its important role in providing information for programmatic learning and improvement, the NACP established evaluation as another central awardee activity.

In 2009, with leadership support and encouragement, the NACP launched an unprecedented strategy to build and advance evaluation capacity among the funded state asthma programs. During a 5-year funding cycle (2009–2014), the NACP designated program evaluation as an explicit priority (CDC, 2009). Managers and evaluators within this national program developed a multi-pronged approach to build evaluation capacity within state asthma programs drawing heavily upon the approach articulated in CDC’s Framework for Program Evaluation in Public Health (CDC, 1999). Central to this approach were requirements that state asthma programs hire an evaluator (at least half-time), create and submit evaluation plans, and routinely report status updates on the evaluative efforts to the NACP. To promote the successful implementation of these new requirements, and to model the importance of evaluation, NACP established their own team of evaluators—the evaluation technical advisors (ETAs). The primary role of the ETAs was to work with evaluators in funded state programs to strengthen and expand evaluation capacity with the ultimate goal of fostering high-quality evaluations that produce findings that stakeholders use.

An internal assessment and an independent study examined the presence of evaluation capacity within state asthma programs operating under the 2009–2014 funding cycle (Evaluation Technical Advisors, 2012; Fierro, 2012). These efforts provided indications that evaluation capacity does exist within these programs—including favorable attitudes and practices toward evaluation for program staff and leadership overall. This evidence, coupled with the intentional efforts performed by the NACP to build evaluation capacity, suggests that evaluative thinking exists among state asthma program staff. Preliminary evidence from conversations and existing data sources suggests that evaluative thinking emerged within some asthma programs. For example, during the 2009–2014 funding cycle the ETAs and several state evaluators noticed subtle changes in the conversations about evaluation—state evaluators described larger, more active roles in program operations and on occasion, others, such as asthma program managers, began to discuss and refer to evaluation as a shared activity. Furthermore, applications for the 2014–2019 funding cycle submitted by existing state asthma programs (CDC, 2014) provide evidence that for many, thinking had transformed. Given this preliminary evidence of evaluative thinking, we embarked on a concerted effort to begin to answer: How do instances of evaluative thinking present in a practice-based setting?

Methods

We performed a multisite case study (Creswell, 2013). Four of the twenty-three states funded under the 2014–2019 cycle by the NACP were purposively sampled to participate; with an emphasis on selecting states that had high levels of evaluation capacity relative to other NACP funded states *and* appeared to have relatively high evaluation capacity over an extended time period. Three sources of data were leveraged for this research: (1) a subset of existing grant administration documents in the NACP, (2) telephone interviews with the evaluator and program manager working on behalf of four state asthma control programs funded by NACP ($N=8$), and (3) a focus group with the same eight individuals. Given the exploratory nature of this research and the limited existing literature on evaluative thinking, we primarily used an inductive approach for our analysis. We describe our methodology in greater detail below.

Sample

The primary inclusion criterion of interest was a high level of evaluation capacity relative to other state grantees. The first step in selecting these states included an assessment of evaluation capacity by the ETAs. Each ETA assigned a value from one to five for the overall evaluation capacity of each funded state program she supported (high from the outset and remained so [5 points]; steadily increased [4 points]; fluctuated over time and topic [3 points]; done as compliance [2 points]; nonexistent [1 point]). Next the ETA used a 5-point frequency scale, where a five represented “routinely exhibits” and a one represented “never/almost never,” to score how routinely each state they specifically work with did or did not exhibit each of the following characteristics: (1) accounting for and engaging an array of perspectives; (2) using evaluation findings; an organizational culture supportive of evaluation, including motivated staff, an evaluation champion, and resources devoted to evaluation; (3) asking meaningful evaluation questions; (4) employing a variety of study methods; and (5) supporting staff ECB. To promote consistent use of the scales, the ETAs discussed potential responses. Since ETA state assignments had changed over time and ETAs regularly have contact with state evaluators, several ETAs weighed in on responses and reached agreement. The scores provided by the ETAs on each of the aforementioned scales were summed for each of the twenty-three states currently funded by the NACP. The eight states with the highest scores were retained for possible inclusion.

Next, a total evaluation capacity score was calculated leveraging data from a previous study (Fierro, 2012) for nineteen of the twenty-three funded states. Evaluation capacity scores for states in which two respondents provided an assessment of evaluation capacity were averaged for a combined state score. One of the eight states rated highest on evaluation capacity by the ETAs did not provide data for the previous study. Of the eight states retained from the ETA assessment described above, the four states with the highest evaluation capacity score from the previous study were selected for inclusion and invited to participate in the current study. The single state missing a value on evaluation capacity from the previous study was tied with another state for fifth place, thus the missing data did not affect the selection. Both the evaluators and program managers from the four selected states had to agree to participate in the study to be included. All invitees agreed to participate. The study was determined exempt by the Institutional Review Board at Claremont Graduate University.

Data Collection

We used three data sources to identify and describe potential instances of evaluative thinking in practice. Data collection (and the associated analyses) occurred sequentially. The first step in this process was a review and analysis of existing grantee documents. Following content analyses of these documents, the team performed semistructured telephone interviews with the evaluator and program manager for each of the four states. Subsequently, our team conducted a telephone focus group with all eight respondents. Each data source is described in additional detail below.

Documents—The sampling strategy we used for our review of archival data was purposive. We selected three documents to review for each state: the state’s most recent funding application; their strategic evaluation plan, a document designed to guide evaluation

efforts over the 5-year grant cycle; and one plan that describes a specific evaluation the state intended to perform. Our team identified these documents in consultation with the ETAs. The ETAs recommended these documents since they were designed explicitly to elicit information from the grantees about evaluation activities occurring in the state. Thus, out of all documents provided by the grantees to the CDC NACP, these were thought to be the most likely to include instances of evaluative thinking. In addition, these three documents were identified by the research team as the most relevant to the study given early reviews of a broader set of documents (described later under Data Analysis).

Interviews—We conducted semi-structured telephone interviews with the evaluator and program manager in each of the four states. All individuals invited to participate in an interview agreed. Interviews lasted approximately one hour and were digitally recorded. Topics included a brief history of evaluation within the state asthma control program, description of the dialogue that has occurred among stakeholders during the program’s evaluation processes, and how (if at all) the interviewee and other evaluation stakeholders’ understanding of evaluation and engagement in the act of evaluation have changed over previous years. The interview protocol was refined prior to administration to ensure that the questions posed were likely to facilitate dialogue about concepts we had identified as potentially important to evaluative thinking which were not seen with a high frequency (or at all) during the document review. The project team created near-verbatim transcripts from these recordings for use in data analysis.

Focus Group—The final data collection effort was a two-hour telephone focus group with all eight study participants. The focus group questions built upon the findings from the interviews with the intention of providing a forum in which participants could reflect on the interview themes related to evaluative thinking and to share their thoughts on existing definitions of evaluative thinking. The group discussion was digitally recorded and transcribed.

Data Analysis

The analysis of data consisted of four sequential steps. First, all team members contributed to developing initial indicators of evaluative thinking. These indicators emerged through a review of a subset of grant documents and were refined through in-depth discussions among our team. Two team members, one with extensive familiarity with evaluation in the NACP and one performing scholarship in the area of evaluative thinking more broadly, used the initial set of indicators to identify instances of evaluative thinking in documents, interviews, and focus groups while continuing to refine the indicators. A description of the process is provided in Figure 3.1.

A Priori Indicator Development—To develop the initial set of evaluative thinking indicators, the full research team reviewed four archival documents, three of which overlapped with the documents purposively selected for the final analyses. The additional document included an interim progress report provided by the grantees as part of an earlier funding cycle (2009–2014). During this initial review of documents, each team member took notes and identified indicators that they believed had relevance given the two existing

definitions of evaluative thinking (Buckley et al., 2015; Vo, 2013) or based upon their experience of what it means to have evaluative thinking in practice. Each team member included a definition and an example from one or more documents for each indicator they identified. In this initial pass through the data, we reviewed documents from six states; each reviewer started the review process with a different state.

Over the course of seven meetings, the research team revised the initial indicator list to facilitate stronger conceptual clarity. During these meetings, we reviewed each proposed indicator, identified overlaps, and extensively discussed potential definitions. From an initial list of approximately seventy possible indicators, we generated a list of sixteen indicators grouped into five overarching categories.

Content Analysis—Document Review—Two team members independently performed descriptive coding for each document using the final set of indicators from the process previously described while allowing for additional indicators of evaluative thinking to emerge. The team coded data in ATLAS.ti 7. Each coder selected illustrative quotes for each indicator, rating them as excellent (full expression of or a tangible example of the indicator), good (indicative of most of the described indicator), or interesting (moderately indicative of the indicator and may offer new insights). The two coders then merged the coded documents, and discussed and reconciled differences. Not all differences were reconciled, for example, in many cases, selected passages were coded by one coder and not another—such differences were retained. Five new indicators emerged from the analysis. The two coders subsequently used thematic analysis to identify common patterns within and across codes, and made final revisions to the indicator list. Final revisions to the indicator list were discussed among the coders during two teleconferences (one of which included the principal investigator) and through e-mail correspondence.

Content Analysis—Interviews and Focus Group—The two members of the team who performed the coding of the existing documents also coded the interview and focus group transcripts. Analysis of the interviews and focus group were performed in sequence (first the interviews, followed by the focus group). The coders independently assigned codes using the indicator list finalized through the document review. The coders merged the coded transcripts and then discussed and reconciled differences. Similar to the analytic procedures used for the document review, not all differences were reconciled and coders allowed additional indicators to emerge. No new indicators emerged from the analysis of interviews or the focus group. The two coders then used thematic analysis to identify common patterns within and across indicators. This analysis was captured in written reflections to inform the study results.

Results

Four state asthma control programs participated in this study—Massachusetts, Montana, Oregon, and Wisconsin. These programs range in the number of years they have received funding under the NACP—from 17 years (Oregon) to seven years (Montana). Two of the four states (Massachusetts and Wisconsin) have external evaluators who have been with the program for several years (approximately seven and four years, respectively), the other two

(Oregon and Montana) have internal evaluators with a tenure less than five years (approximately four and two and a half years, respectively). Program managers have had varying tenures with their respective programs with the manager in Wisconsin coming on board approximately 11 years ago and the Oregon manager joining three months prior to the interview.

Several of these states have unique contextual features worthy of mention as they may affect the dynamics of the evaluation culture within their program. The presence of such contextual features may contribute to greater ease or receptivity in engaging in ECB efforts compared to other programs or to higher baseline level of evaluative thinking prior to engaging in ECB efforts. In Montana, the program manager served as the epidemiologist and evaluator for this same program for approximately five years prior to switching roles. She now serves as the manager of the asthma program as well as the supervisor for a larger section that includes several public health programs. In Oregon, the state asthma control program is not a stand-alone program, rather it is deeply integrated with several other chronic disease programs in their Health Promotion and Chronic Disease Prevention Section. Evaluators in this section are centralized—they work across disease-specific programs but do have an emphasis area (e.g., asthma). And in Wisconsin, the current evaluator has an interrupted tenure with the program—serving as the evaluator for approximately two and a half years near the initiation of the requirement for evaluation and subsequently returning about one year prior to when we engaged her in an interview.

Indicators and Examples of Evaluative Thinking in Practice

Twenty-two indicators of evaluative thinking grouped into five categories emerged through our analyses (Figure 3.2)—reflecting, perspectives, projecting, valuing evaluation, and use. In this section, we present each indicator of evaluative thinking that emerged from the data, share the definition we developed for the indicator, and provide an example of how the indicator manifested in practice by providing text from a document, interview, or the focus group (Table 3.1). In addition, we share some general reflections on the frequency with which the indicators were identified across different types of data sources.

Reflecting—We defined reflecting as “deliberately giving critical attention to various aspects of a program, including its context and its evaluation; suggests a willingness to apply a critical lens reflexively.” All six indicators under this category were identified a priori—four of these indicators emerged based upon our understanding of existing scholarship on evaluative thinking (thoughtful questions, deeper understanding, describing thinking, and identifying assumptions) and two based upon our experience as practitioners (considering context and evaluation review). Of these indicators, text was assigned most frequently to considering context ($n = 53$ instances across all data sources). Coders infrequently identified text that aligned with the indicators of identifying assumptions ($n = 3$ instances in interviews only) and evaluation review ($n = 3$ instances across documents and interviews).

Perspectives—Indicators under this category relate to “incorporating information and priorities from multiple viewpoints.” Four indicators of evaluative thinking were identified under this heading, three of which were defined a priori and informed by our work as

practitioners (multiple perspectives, additional points of view, and explicating values) and one which emerged during the formal content analysis of documents (participatory evaluation). Multiple perspectives and participatory evaluation (an indicator of statements expressing the value of participatory evaluation) were the indicators most frequently identified by the coders across data sources ($n = 32$ and $n = 19$, respectively) despite the later not being an a priori code. Multiple perspectives, an indicator that suggests the grantee incorporated different points of view, was infrequently identified by coders ($n = 1$ instance in an interview); it was however seen more frequently in documents ($n = 7$).

Projecting—We defined projecting as “envisioning success and the path to achieving it.” Of the four indicators categorized under this heading, two were identified a priori in light of our professional experience (criteria of success and linking activities to outcomes) and two emerged during the formal analysis of documents (scaling and suite of evaluation activities). All of the indicators under this category were identified much more frequently in the document review compared to the interviews and focus group despite two not being identified a priori. Coders assigned one of these four indicators to 118 instances of text in the documents reviewed and in only four instances of text in transcripts from interviews and the focus group. Only one instance of text (in a document) was associated with the emergent indicator suite of evaluation activities (an indicator that evaluation activities are performed as a suite of studies as opposed to individual events). Despite the limited identification of this indicator, we retained it due to its potential to reflect evaluative thinking in future studies.

Valuing Evaluation—Indicators included in this category indicate a belief in the importance and utility of evaluation. All of these indicators were identified a priori—two were created in alignment with our understanding of existing theory related to evaluative thinking (value of evaluation and value of evidence), the other two (intent to engage in ECB and distributed responsibility) were identified as potential evaluative thinking indicators as a result of our practice-based experiences and knowledge of the ECB literature. Text associated with value of evidence (which included statements suggesting a belief in the value of evidence generally) and intent to engage in ECB was most frequently identified in the documents ($n = 155$ instances and $n = 25$ instances, respectively). Value of evaluation (statements suggesting a belief in the value of evaluation) and distributed responsibility (statements indicating that people across the program are responsible for conducting and using evaluations) were the most frequent indicators identified in interview and focus group text ($n = 53$ and $n = 24$, respectively).

Use—We defined use as “the impact or intended impact of the evaluation on the evaluand, stakeholders, and/or society.” We identified four indicators under this category—two of which were identified a priori (planning for use and instrumental use) and two of which emerged during the content analysis of documents (integration and process use). The study team leveraged existing theory and literature related to evaluative thinking to inform the creation of each indicator under this category. The two a priori codes were assigned with differing frequencies across data sources. Planning for use was more often detected in the document review ($n = 74$) compared to the interviews and focus group ($n = 6$) whereas

instrumental use was assigned with almost equal frequency across these data sources ($n = 21$ instances in document review, $n = 16$ in interviews/focus group).

Every program had evidence of at least one indicator from each of the five categories. The indicators assigned least frequently to text from any data source included suite of evaluation activities ($n = 1$), evaluation review ($n = 2$), and identifying assumptions ($n = 3$). The indicators assigned most frequently across all data sources included value of evidence ($n = 163$) and planning for use ($n = 80$). Six indicators were represented by all four states—(1) describing thinking, (2) considering context, (3) participatory evaluation, (4) value of evaluation, (5) distributed responsibility, and (6) instrumental use. All states missed at least one indicator. The number of indicators not identified in text associated with a specific state ranged from one to five indicators.

Patterns in Indicators of Evaluative Thinking

In response to our research question—*How do instances of evaluative thinking present in a practice-based setting?*—we found instances of evaluative thinking across all data sources we used (existing grant documents, interviews with program staff, and a focus group with program staff). However, indicators of evaluative thinking appeared with varying frequencies across data sources.

In reviewing the indicators that emerged through this study, we identified one pattern that may be helpful to consider in future efforts. The indicators appeared to fall into three domains. First, some indicators appeared to be direct indicators of evaluative thinking such as posing thoughtful questions or illustrating thinking. A second set of indicators are potentially supportive of or help to facilitate evaluative thinking—such as additional points of view which reflects that grantees considered views and perspectives on the evaluand beyond the stakeholders who they directly engaged in the evaluation. Finally, there were other indicators that appear to be a behavioral manifestation of one or more individuals thinking evaluatively within a program such as the intent of the grantee to engage in ECB, as evidenced through documented plans to build capacity, or distributing responsibility for evaluation throughout their program.

In addition to the three domains of indicators we just described, it became apparent that the indicators we identified had some sort of relation to each other. Figure 3.3 presents one of several potential conceptualizations of how the evaluative thinking indicators we identified relate to each other based upon our interpretation of the data. The five categories, and their associated indicators, represent critical junctions in evaluative thinking. A key intersection is located at the center of the figure, where projecting, perspectives, and valuing evaluation meet to facilitate reflecting. According to the data, the act of reflecting often occurs in the form of dialogue, where groups explore complex issues from various viewpoints and values to discover new knowledge and understanding. From this, evaluative arguments and desired actions (i.e., use) become apparent.

We also realized the patterns of indicators manifested collectively among the program staff. Evaluation was not seen as the sole responsibility of the “evaluator”—indeed the value of communications among stakeholders became a common practice. Repeatedly, we heard that

evaluative thinking was a shared concept, with evaluation being a language and a process that the program staff learned together. This intentionality permeated a wide variety of program activities beyond evaluation itself (e.g., strategic planning of programs, implementing programs).

Strengths and Limitations

Several strengths were present in our study procedures. Perhaps foremost is the diversity of perspectives represented on our team. As previously noted, our team was comprised of several individuals with extensive practice-based expertise developing and implementing ECB efforts with the NACP and beyond as well as several individuals who were deeply embedded in scholarly activities that examine evaluative thinking and the closely related topic of ECB. In addition, half of our team members were deeply familiar with the NACP and the specific states selected for this study whereas this was the first exposure the other half of our team members had to this program. It is likely that this mixture of experiences helped us to identify a comprehensive set of indicators of evaluative thinking.

This comprehensiveness, however, may have contributed to some of the limitations of our study as well. Some indicators that our team surfaced, particularly those identified a priori, were never identified in a specific data source. This could be a result of several factors in addition to developing an overly comprehensive listing of indicators. For example, it may be the case that some indicators are just more likely to be apparent in certain data sources.

Our interrater reliability (IRR), or rather lack thereof, for some indicators could also be viewed as a limitation. Using the findings from the final content analyses, IRR ranged from a Cohen's kappa of -0.594 (moderate disagreement) to 0.99 (almost perfect agreement). However, it is important to note that only four indicators reported kappa values suggesting substantial or almost perfect agreement (Landis & Koch, 1977). Such differences, however, are not unanticipated due to the different backgrounds and experiences of the two coders. Ultimately, we do not view this lack of agreement as a limitation, given that this was an exploratory study designed to acquire a first glance of how evaluative thinking might manifest in practice. In fact, one might view a lack of agreement as positive and anticipated in such an endeavor—positive in that it leads to a more comprehensive listing of potential instances of evaluative thinking and expected since we intentionally engaged team members with diverse perspectives for the purpose of facilitating a broad representation of evaluative thinking.

As in any research endeavor, there are aspects of this inquiry process that we would refine in retrospect. During the writing phase, we frequently lamented that we did not have the person-hours available to go back through each data source (particularly the interviews and focus group) to further refine the codes, indicators, and associated definitions. Another round of coding, reflection, and refinement would have afforded us the opportunity to develop a set of indicators that is more parsimonious and defined more specifically than presented in this manuscript. For example, during the final reviews of this manuscript a guest editor posed a question about the definition for “deeper understanding”—specifically whether it was possible to state who “seeks insights and probes for deeper understanding.”

At this juncture, we do not have that level of specificity—statements without specific actors that pointed to seeking insights or deeper understanding were included under this code. However, a second or third iteration of coding may have resulted in a clearer interpretation of “who” exactly is involved and/or whether statements absent such specificity should be retained under this code.

Another round of analysis would have also provided our team with an opportunity to examine a theme that we suspect is important to evaluative thinking, but were unable to pursue in depth. We noticed frequent mention/discussion of “conversations” in the interviews and focus group. It seems that the idea of conversations may lie at the intersection of several indicators we presented in this manuscript (e.g., “describing thinking” and “identifying assumptions” from the reflecting category and “multiple perspectives” and/or “explicating values” from the perspectives category). If we had coded “conversations” as a mechanism to get at one way people share their perspectives, important insights may have emerged that we are not able to share at this time.

Such reflection points to another component of this research process that could have benefited from refinement. The process of defining and refining indicators was incredibly challenging, in fact we often noted that we felt as though we were “wrapping ourselves up in cognitive pretzels.” Perhaps this was due to having to reconcile the multiple, diverse perspectives represented by our team members, perhaps it is simply just a natural part of this process. Irrespective, it may have been helpful to seek reviews of our indicators and definitions from others who have an interest in the topic of evaluative thinking. Such reviews could have helped to show us where further refinements or clarification in our articulation of these indicators would be helpful for the intended end-users.

Discussion

Our goal in this research was to move the concept of evaluative thinking from the intangible to the tangible. We hope this research provides insights to practitioners about what they may wish to look for over time and the types of data sources that may be more or less helpful in making observations about the existence (or growth) of evaluative thinking. Certainly, the indicators we developed as part of this exercise can benefit greatly from refinement in future research on evaluation efforts and we encourage others to be critical of what we have provided here and to work to improve and expand upon this effort.

Some suggestions for future research include performing one or more longitudinal studies that examine the emergence of evaluative thinking. Such studies might examine the extent to which evaluative thinking (and specifically what aspects of it) are associated with, if any, different ECB approaches. For example, what aspects of evaluative thinking are most often associated with participatory evaluation approaches that may intentionally seek to build evaluation capacity through the mechanism of process use and how does this differ from other ECB approaches such as multicomponent approaches offering a suite of training, one-on-one technical assistance, and guidance documents from a funding entity?

Further, future research might explore the relationships between different types of indicators of evaluative thinking as depicted in Figure 3.3 or the unit of analysis most appropriate for identifying evaluative thinking (i.e., individual-level change or group-level changes). Finally, additional research similar to the current study conducted in different contexts can help us to understand the extent to which what we saw extends to other contexts and may be quite helpful for practitioners who hope to continuously reflect on the extent to which they can foster and are realizing evaluative thinking in their settings.

In structuring future research, particularly studies including participants similar to ours, it may be important to consider the extent to which indicators of evaluative thinking are reflective of evaluative thinking on the part of grantees versus evaluative thinking on the part of the funder. The guidance provided by NACP regarding how to perform evaluation within funded programs frequently (but not always) aligned directly with text to which we assigned an indicator. For example, the NACP's approach to evaluation aligns with CDC's Framework for Program Evaluation in Public Health which draws upon a Utilization-Focused Evaluation approach (Patton, 2008), emphasizes the importance of stakeholder engagement in evaluations, and stresses the importance of developing a program theory (CDC, 1999). With this backdrop, perhaps it is not surprising that illustrating thinking, linking activities to outcomes, and planning for use were three of the four most frequently identified indicators in the content analysis.

Similarly, "value of evidence" was by far the most frequently identified indicator ($n = 166$ instances across all data sources). One of the documents included in the content analysis was the grantee's application for the most recent round of funding (2014–2019), which was titled "Comprehensive Asthma Control Through Evidence-Based Strategies and Public Health–Healthcare Collaboration." Not surprisingly, language indicating that the state program intended to implement evidence-based strategies was used repeatedly throughout the application.

We are not suggesting that NACP's evaluation guidance is solely responsible for the emergence of the evaluative thinking indicators we identified. After all, some indicators emphasized extensively in the guidance from the NACP were not encountered in data sources from each of the four states such as "intent to engage in ECB." What we suspect is that indicators of evaluative thinking arise from a complex interplay between individuals in a system and that some portion of this is influenced by federal evaluation policies and practices. Future research could examine such interplays.

Ultimately, our study suggests that funders do have the potential to affect evaluative thinking through the activities they prescribe as part of their ECB efforts. In this case, some of the ECB efforts that may have contributed to evaluative thinking within the participating states include:

- Asking questions that require thoughtful reflection,
- Requiring information about the level and type of engagement of multiple stakeholders as part of the evaluative processes,

- Promoting use of different types of evaluation designs and methods throughout a program's lifecycle,
- Requiring reporting of ECB activities,
- Promoting visual depictions of how activities link to outcomes, and
- Requiring reporting of how evaluation findings were used.

Reflection of the extent to which evaluative thinking has emerged in programs with similar concerted ECB efforts could help to expand our collective understanding of evaluative thinking and the conditions that foster its development.

References

- Alkin, MC. Evaluation roots: A wider perspective of theorists' views and influences. 2. Thousand Oaks, CA: Sage; 2013.
- Bruner Foundation, Inc. Evaluative thinking assessment tool. 2010. Retrieved from <https://www.evaluativethinking.org/evalthink.html>
- Buckley J, Archibald T, Hargraves M, Trochim WM. Defining and teaching evaluative thinking: Insights from research on critical thinking. *American Journal of Evaluation*. 2015; 36(3):375–388.
- Centers for Disease Control and Prevention. Framework for program evaluation in public health. *MMWR*. 1999; 48(RR-11)
- Centers for Disease Control and Prevention. Addressing asthma from a public health perspective. 2009. Retrieved from https://www.dttac.org/services/DPCP_101/pdfs/FOA_DP09-901_FOA_FINAL.pdf
- Centers for Disease Control and Prevention. Comprehensive asthma control through evidence-based strategies and public health and health care collaboration. 2014. Retrieved from <https://www.grants.gov/web/grants/view-opportunity.html?oppId=252419>
- Cornell Office for Research on Evaluation. Evaluative thinking downloads. n.d. Retrieved from https://core.human.cornell.edu/documents/ET_Measures.cfm
- Creswell, JW. Qualitative inquiry and research design. 3. Thousand Oaks, CA: Sage; 2013.
- Earl, L., Timperley, H. Evaluative thinking for successful educational innovation. Paris: OECD Publishing; 2015. (OECB Education Working Papers 122)
- Evaluation Technical Advisors. Evaluation of the provision of evaluation technical assistance to states funded by the National Asthma Control Program (Internal document). 2012 Feb. Available on request from mwilce@cdc.gov.
- Fierro, LA. Clarifying the connections: Evaluation capacity and intended outcomes. The Claremont Graduate University; 2012.
- Greene JC. Stakeholder participation and utilization in program evaluation. *Evaluation Review*. 1988; 12(2):91–116.
- King, JA. Developing evaluation capacity through process use. In: Cousins, JB., editor. *Process use in theory, research, and practice*. New Directions for Evaluation. Vol. 116. San Francisco, CA: Jossey-Bass; 2007. p. 45-59.
- Labin SN, Duffy JL, Meyers DC, Wandersman A, Lesesne CA. A research synthesis of the evaluation capacity building literature. *American Journal of Evaluation*. 2012; 33(3):307–338.
- Landis RJ, Koch GG. The measurement of observer agreement for categorical data. *Biometrics*. 1977; 33:159–174. [PubMed: 843571]
- Patton, MQ. Utilization-focused evaluation. 4. Thousand Oaks, CA: Sage; 2008.
- Patton, MQ. *Essentials of utilization-focused evaluation*. Thousand Oaks, CA: Sage; 2012.
- Rich, RF. Uses of social science information by federal bureaucrats: Knowledge for action versus knowledge for understanding. In: Weiss, CH., editor. *Using social research in public policy making*. Lexington, MA: Lexington Books; 1977.

- Russ-Eft, D., Preskill, H. Evaluation in organizations: A systematic approach to enhancing learning, performance, and change. 2. New York: Basic Books; 2009.
- Shulha LM, Cousins JB. Evaluation use: Theory, research, and practice since 1986. *Evaluation Practice*. 1997; 18(3):195–208.
- State, Tribal, Local, & Territorial Public Health Professionals Gateway. Public health system and the 10 essential public health services. n.d. Retrieved from <https://www.cdc.gov/stltpublichealth/publichealthservices/essentialhealthservices.html>
- Vo, A. Unpublished doctoral dissertation. Los Angeles, CA: University of California; 2013. Toward a definition of evaluative thinking.
- Weiss, CH. Evaluation. 2. Upper Saddle River, NJ: Prentice Hall; 1998.

Biographies

Leslie A. Fierro is assistant clinical professor of evaluation at Claremont Graduate University. She previously worked as an evaluation technical advisor in CDC's National Asthma Control Program.

Heather Codd is a doctoral student in evaluation and applied research methods at Claremont Graduate University. Her current research interests include organizational learning, systems thinking, and evaluative thinking.

Sarah Gill is a behavioral scientist in CDC's Program Performance and Evaluation Office. She was formerly an evaluation technical advisor in CDC's National Asthma Control Program.

Phung K. Pham is a doctoral student in evaluation and applied research methods at Claremont Graduate University. She is also a data analyst for the Division of Emergency and Transport Medicine at Children's Hospital Los Angeles.

Piper T. Grandjean TARGOS is a doctoral student in evaluation and applied research methods at Claremont Graduate University. She enjoys working with organizations to build their capacity to engage in effective evaluations.

Maureen Wilce is an evaluation technical advisor in CDC's National Asthma Control Program which is located in the Asthma and Community Health Branch, Division of Environmental Health Science and Practice, National Center for Environmental Health.

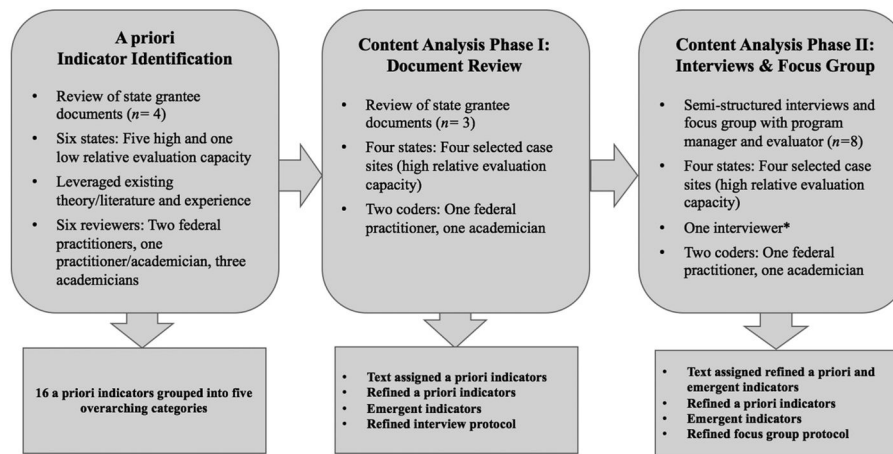


Figure 3.1.
Description of analysis process

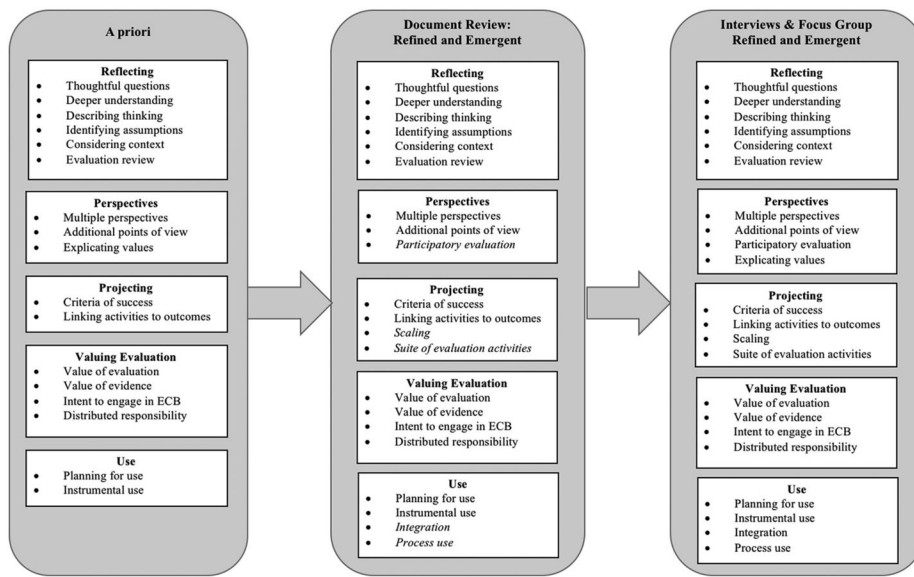


Figure 3.2.
Identification of evaluative thinking indicators

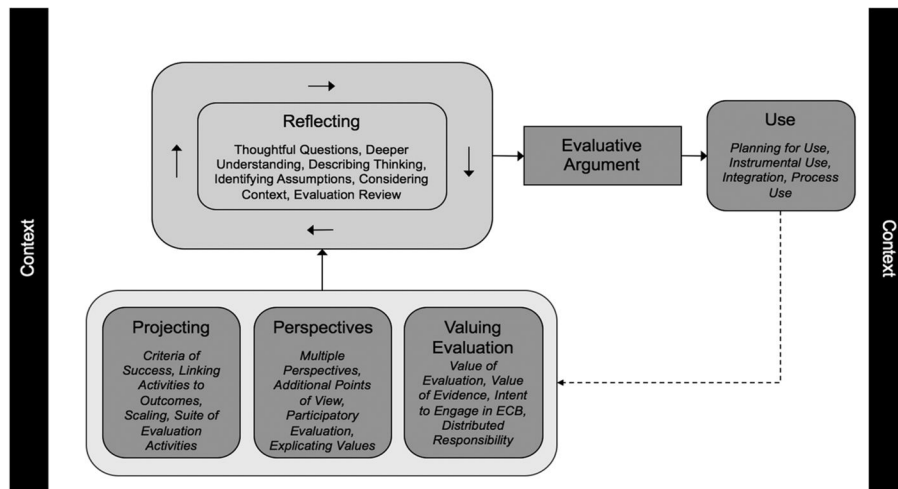


Figure 3.3. Conceptual model of potential relationships among evaluative thinking indicators

Table 3.1

Indicators of Evaluative Thinking

Indicator	Definition	Example (Source)
Reflecting—Deliberately giving critical attention to various aspects of a program, including its context and its evaluation; suggests a willingness to apply a critical lens reflexively		
Thoughtful questions	Organizational environment/individual attitudes reflect a stance of “fearless inquiry.” Asking questions that look at the big picture; suggest an openness to discovery and a willingness to encounter negative findings and unintended consequences; examines deeply held beliefs.	“The first thing that comes to mind is being curious and asking questions about the programs we coordinate, supervise, or administer. Being thoughtful about what works well, what doesn’t work well, are we serving our customers. Being thoughtful about what we are doing and why and being willing to ask these questions.” (Focus group).
Deeper understanding	Ongoing tendency to seek insights and probe for further information and explanation about the evaluation findings.	Some evaluation results led to the development of future evaluation questions like further investigation into why families discontinue the MAP (Document—Application).
Describing thinking	Communicating ideas and concepts as well as the logic or thought processes on which they are based, including describing or defining words or situations in one’s own words, or illustrating a concept visually (e.g., logic model, TOC). This activity is both a learning tool (working out the connections, ideally with others) as well as an indication of internal (personal) comprehension (Buckley et al., 2015) and communication device.	“we’ve kind of tried to figure out things, you know, how they’re prioritized and then think of things, you know, these small little things that—we can’t evaluate everything and that could be someone who could just do you know a very brief kind of evaluation, and then, kind of, what needs to be saved more for, kind of, more in-depth evaluation.” (Interview).
Identifying assumptions	Identifying and articulating underlying concepts and beliefs.	“We were so focused on trying to get, finding solutions, that we were asking the wrong questions” (Interview).
Considering context	Demonstrating awareness and responsiveness to the prevailing context and to changes that are occurring in their environment (Earl & Timperley, 2015). Program and evaluation designs are flexible and adaptive.	“People are really curious about the results and they like to give feedback, but to have them at the table from the start of the planning process is not something that’s super feasible that we’ve found with the people we’ve worked with because we do work with a lot of healthcare providers.” (Interview).
Evaluation review	Program makes efforts to examine or judge the quality of its evaluation work and/or assess the performance of evaluators.	The OAP will document evaluation lessons learned in the course of implementing the strategic evaluation plan by incorporating a meta-evaluation component in the annual Strategic Evaluation Review process. OAP will review the changes made to the strategic evaluation plan and the reasons for the changes (Document).
Perspectives—Incorporating information and priorities from multiple viewpoints		
Multiple perspectives	Soliciting a diverse range of stakeholder views and perspectives on the evaluation.	“We focus a lot on having the users of the evaluation, who would benefit the most from the evaluation, be a part of the entire process. Engaging these different users throughout the process of the evaluation, collecting the data, using the results. We are very adamant on making sure these folks are involved in the conversations and have input on what would be most important for them to learn from the evaluation and what evaluation questions they want to answer so we know that these evaluations aren’t just for the state program but the partners and stakeholder to use. It’s important that they are involved so they can identify what needs to change and what will make programs sustainable moving forward” (Focus group).
Additional points of view	Consideration of views and perspectives on the evaluation beyond stakeholders who can be directly engaged.	“So, the one we haven’t as much directly is children themselves or youth themselves. They haven’t necessarily been as much the focus of our design, partially because HIPAA and the confidentiality of actual clients. So we talked more to parents who are advocates in local coalitions and community health workers who work with clients but we haven’t gone directly to patients and clients” (Interview).

Indicator	Definition	Example (Source)
Participatory evaluation	Declarations of the value of/commitment to broad participation in evaluation activities beyond the evaluator and program leadership.	“So, because we look at evaluation as being useful to, to those that are actually taking part in the program and that, um, are most affected, which a lot of the time it isn’t us but we’re the ones facilitating the activity or the program, and so, um, it’s a user driven evaluation, um, model that we want to make sure that everyone’s feedback and input, um, that are actually impacted by the program and activities, um, has had an opportunity to contribute to the evaluation because they were the ones that experienced firsthand um, the process, and um, will care as much as we do, if not more, about the outcomes” (Interview).
Explicating values	Consideration of the values that are pertinent to the evaluation and can/should influence the evaluation process, the program theory (e.g., outcomes), what constitutes credible evidence, and the findings/ recommendations.	“We learned that, that really showcased for us the value of the stakeholder-driven model to improve and help in focusing evaluation candidates and setting the stage for individual evaluation design and questions that were key to stakeholders and was as going to be valued by stakeholders and perceived as strong evidence to shape, to meet their needs and shape some of their decision-making.” (Interview)
Projecting—Envisioning success and the path to achieving it		
Criteria of success	Clearly articulating criteria for success.	The project will work with six PWTG grantees in ten school districts, reaching an estimated 1333 students with asthma (in Year 1), to integrate clinical activities with school asthma control activities; reduce environmental triggers in schools; Improve the self-management education of children with asthma and their families and improve the nurse management of high-risk asthma (Documents—Application).
Linking activities to outcomes	Articulating the intended/expected connections among various elements of the program theory or program implementation.	Additionally, OHA-PHD will provide training and technical assistance to clinical providers on guideline-based care for asthma through Housing with Services, a demonstration project to enhance clinical care and reduce hospital and emergency room use among low-income seniors (Documents—Application).
Scaling	Planning for the ultimate outcomes of a program, as with references to state-wide or population-level outcomes or program/outcome sustainability.	The requirements to scale home visiting are capacity to deliver a high-quality intervention and insurance reimbursement for the service. These variables are linked: large-scale capacity development requires reimbursement, but payers will be reluctant to reimburse for a service delivered outside the clinic walls, even if it is evidence based, unless they can be assured that the intervention is delivered with fidelity to the evidence-based model (Documents—Application).
Suite of evaluation activities	Evaluation activities are conducted as a suite of studies as opposed to individual events. While each evaluation has its unique focus, the intent is to gradually bring clarity to the bigger picture (e.g., pieces of a puzzle).	During the last funding cycle, the evaluation focused on infrastructure strategies including strategic partnerships (WAC, surveillance (Burden of Asthma), and evaluation. Each of these was shown to have significant strengths and sustainability. Therefore, the evaluation team reported that they would prefer to focus on the services and health systems strategies for the evaluation during this funding cycle (Documents—Individual Evaluation Plan).
Valuing Evaluation—Statements indicating belief in the importance and utility of evaluation		
Value of evaluation	Statements indicating belief in the importance and utility of evaluation.	“If we dream as big as we can and then get down to the nuances of the day-to-day, I think for us the evaluation piece is so important to making that all possible—it allows us to sit back and be objective about what we are doing and the scan of our program, be it the epi or our program staff or the evaluation staff. It gives us the tools and the language and the framework to talk about how we get from point a to b and make sure we are meeting our goals, serving folks in the most effective ways we can, helps us to be conscience of limited resources. It seems like a guiding principle and it’s integrated into everything we do that it just makes sense and gives us the framework for moving forward and has become part of the guiding philosophy of the program rather than just a one-off evaluation here and there, not that that doesn’t happen but it definitely happens in conjunction within the overall expectation that our evaluation work is one of the core tenants of what we do and will be at the forefront” (Focus group).
Value of evidence	Judgments, values, and assertions about the program are informed by findings attained through systematic inquiry and credible methodology. When such findings do not exist, efforts are made to attain them.	“So I feel driven to make sure that we are maximizing the resources we have to have the greatest impact. To me that means, if we are piloting something new, it should be a real pilot. We should find out, does this work. I want us to not be driven by the “gee this sounds good.” And this probably comes from my research background, wanting to follow the evidence and follow the whole experience and make sure we’re doing things that are well rooted in, and give a true view as we can have of the current situation, the current world. That really drives this desire to constantly be striving for that data, that information that can then help us have stronger programming, greater reach, more effectiveness with the limited resources that we have.” (Interview).

Indicator	Definition	Example (Source)
Intent to engage in ECB	Concrete documented efforts or plans to design and implement teaching and learning strategies to: (1) help individuals, groups, and the organization learn about what constitutes effective, useful, and professional evaluation practice and (2) support sustainable evaluation practice through such things as changes to organizational norms or infrastructure (e.g., evaluation policies, technology) (Russ-Eft & Preskill, 2009).	Over the last 5 years, the program has built the capacity to execute evaluations and act on the results and will continue to learn and act as a team to evaluate future efforts (Infra V.B.1). This workgroup has planned evaluation trainings for the entire CDPHP Bureau and regularly works together as an evaluative learning community to support each other's evaluations. The MACP commits to collaborate on cross-state evaluations and other evaluative learning communities during this funding cycle (Documents—Application).
Distributed responsibility	People in a variety of roles at various levels of implementation are responsible for conducting evaluations and using the findings.	“We have an evaluation consultant who works with us directly, but also have an understanding that everyone on the team should have at least a basic understanding of the evaluation work that we do. There is an expectation that it is embedded into all our program work. Evaluation is part of the comprehensive approach to everything we do” (Focus group).
<i>Use—Impact or intended impact of the evaluation on the evaluatee, stakeholders, and/or society</i>		
Planning for use	Evaluation planning includes explicit consideration for the ways in which evaluation findings will be used to support subsequent thoughts and actions about the program and who the users will be. May also include specific actions taken to lay the groundwork for use (Alkin, 2013; Weiss, 1998).	“We looked at a timeline. We kind of laid out like when is our grant reporting due? The legislature in Montana convenes every 2 years, and so when do we need to have information for that? When do we need to have different fact sheets and things updated? So we laid out when we would need information by” (Interview).
Instrumental use	Instances where the evaluation findings were used to modify the program in some way that improves alignment with program theory, will assist in the achievement of outcomes, and/or will mitigate negative consequences (Greene, 1988; Rich, 1977; Shulha & Cousins, 1997). Frequency may vary; may occur with interim or final findings.	“We did end up pulling the money from that agency and putting it elsewhere, and so again, you know, I think that the huge importance of the evaluation—is the programming working, if it's not let's do something to change, to change it” (Interview).
Integration	Evaluation is embedded into the routine practices of the organization. People in the organization expect to engage with each other in clarifying key concepts, examining the quality of evidence available about effectiveness, and supporting their opinions and judgments with evidence.	“I've really gotten more to the point of trying to integrate evaluation as a systematic process in the strategic planning process, so that it is a continuous integrated system rather than a thing you do, from time to time, when needing to improve something specific. Then I would say that that has been the broadest—not just for me but for others—change over time, is really thinking about it as, as a part of systematic business, rather than something extra that is to be done sometimes” (Interview).
Process use	Instances in which program staff (may have) learned about or made changes to the program as a result of the evaluation process rather than the evaluation findings (Patton, 2012).	“At first the conversations were focusing on the end user but we learned things about our work at a much higher level then and I think that's where we really struggled and we couldn't quite figure it out because of that we had implications at the end user level that was frustrating. But it was really illuminating to figure that out through the evaluation process and to expand our understanding of what went wrong and to think about what we could have done to improve it in the future” (Interview).