Building a Fractal Evaluation:
A Study on Improving Skills and Attitudes towards Evaluation Through a Design Similar to Program Pedagogy

A THESIS
SUBMITTED TO THE FACULTY OF THE
UNIVERSITY OF MINNESOTA
BY

Anika A. Taylor

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR THE DEGREE OF
MASTER OF ARTS

Jean A. King, Advisor

May 2015

Acknowledgments

I would like to acknowledge the hard work and contributions of the NISENet Team-Based Inquiry designers and trainers, Sarah Cohn, Scott Pattison, Liz Kollmann, Ali Jackson, Brad Herring, Scott Randol, and Kevin Dilley whose commitment to sharing and growing the community of evaluation within the informal science community will be making ripples for years to come.

The perspective, guidance, and assistance of many faculty at the University of Minnesota were instrumental in my ability to complete this study. In particular, I would like to thank Jean King for her patient support and insightful feedback during this process, as well as Frances Lawrenz and Amy Grack-Nelson for all their help in developing interviewing skills and strategy.

I would also like to thank the education staff of The Bakken Museum for their willingness and enthusiasm to embark upon this project and for taking the time to share their reflections and insights on the process with me.

Dedication

I dedicate this thesis to my parents, who infected me with academia when I was young, but have kindly nursed me through as it grew worse with age, to my brother who has coached me through this and many other trips down many other rivers, and to the staff of The Bakken Museum, who continue to inspire a passion for science in me as well as in those touched by their programs.

Abstract

This thesis is a study on using an evaluation model that mimics the educational pedagogy of an organization as an approach to evaluation capacity building. Referred to as a fractal evaluation, this model of using a “self-similar” approach to learning in an organization sought to improve in staff both their skills in and attitudes around evaluation, or their Evaluation Assets. After the implementation of this model in the education department of a small science museum, interviews and surveys were conducted with staff to determine the effectiveness of using this approach to improve their attitudes and mindsets around the practice of evaluation. The results of the study suggest that staff attitudes and interest in the topic both improved notably along with an expressed increase in staff investment in the process and desire for ongoing learning about program evaluation.

Table of Contents

List of Tables

List of Figures

List of Abbreviations

ECB—Evaluation Capacity Building

NISENet—Nanoscale Informal Science Education Network

P-PE—Practical-Participatory Evaluation

TBI—Team-Based Inquiry

UFE—Utilization-Focused Evaluation

WTDS—Wonder, Try, Discover Share

CHAPTER 1 --INTRODUCTION

Problem Statement and Purpose of Study

Introduction

For many organizations evaluation holds the role of the intelligent but unpopular student at school. While potentially having a great deal to offer, insights and information abounding, it struggles to be accepted much less welcomed, and when it does get a chance to share its ideas, it is often hampered by an awkwardness of language or comportment. This could be because evaluation often communicates its ideas in a manner that is too theoretical or too practical, too abstract or too blunt, too broad or too narrow, or perhaps the problem is as stark as Weiler and Stearns (1980) asserted in their assessment of the struggles of evaluations in a policy making context, “They [evaluations] often fail to address the right questions; promise more answers than they can deliver; produce reports late in the decision cycle; and measure outcomes too narrowly “ (p. 78). Perhaps, in some cases, it is because evaluation is perceived to be an enemy entity; stakeholders and program staff are often skeptical at best, and antagonistic or dismissive at worst. In these situations, evaluation could benefit from learning to speak the language, or to enact the social cues of its subjects. An evaluation that can, like a student learning how to mimic the social norms of his or her classmates, make itself seem to be endemic to a program rather than contrary to it will likely stand a much better chance of having its insights heard and recognized for the intelligence they bear.

Despite these challenges, evaluation has unequivocally earned a place at the lunch table—if only because the principal requires that it be so. Thus, it becomes even

more crucial for evaluation and its subjects to play well together, as time and resources are being allocated to the endeavor and it would be quite tragic for the results to lack use or consideration. To address these challenges, and to posit a possible solution for a cultural shift, or social integration, of evaluation in an educational context, I present the case of the Education Department at The Bakken Museum.

The Bakken Museum, a museum of electricity and magnetism located in Minneapolis, is a physically small institution housing a burgeoning and active education department with a large presence in schools and school districts across the Twin Cities Metro Area. The education department of The Bakken comprises numerous programs that, in turn, each comprise numerous workshops and activities. The museum's mission, "To inspire a passion for science and its potential for social good by helping people explore the history and nature of electricity and magnetism" (Bakken 2013), is heavily evident at each level of the department, with a particular focus on the inspiration of passion. The education department has developed what they call the Science Assets™, the skills and attitudes necessary to succeed in the scientific world (Murphy, 2012). This set of assets includes ideas ranging from specific creative thinking skills to students' ability to identify themselves as capable and enthusiastic towards scientific endeavors, to the relevance of the scientific field in their lives; the simplest version of the program theory teaches students that doing science is really about solving problems using four simple pieces of a puzzle: Wonder, Try, Discover, and Share. The image of the puzzle is used to promote the non-linear and non-finite, or iterative, nature of these tools; it is The Bakken's own version of the scientific process.

Problem Statement

While some pieces of Bakken education programs have been evaluated due to requirements for specific grants, these have been fairly isolated experiences and the staff's general attitude towards evaluation has been unenthusiastic and somewhat suspicious. Given that the organization is small and faces many of the challenges of other non-profit organizations, including budgetary restrictions and limited timeframes, evaluation has not been a priority. While the culture in the museum is generally a self-critical one, with program staff frequently considering and re-considering their practices, the opportunity for any sort of thorough or intentional study has not been explored. Still, as a non-profit, much of The Bakken's funding is grant-based and requires some sort of demonstration that funds have been used to positive effect or, in the case of grants being sought, that funds would be used effectively and that that effectiveness would be concretely demonstrated in future reports. As stated by Pattison, Cohn, and Kollmann (2013):

> "…over the last several decades, more and more private and governmental agencies are requiring that data collection be an integral part of the projects they fund, both to inform the development and implementation of those projects and to assess their impact and the extent to which they have achieved their intended goals." (p. 5)

The Bakken, like so many other organizations of its ilk, is in need of data to support its endeavors as well as to inform decisions about programs and program development.

The staff of this program come to it from a wide variety of vocations; they are performers, research scientists, classroom teachers etc. This diversity brings a wealth of perspectives on developing and refining the art of inspiring the public to a life of interest in science. It also brings a host of skepticisms that are not all easily addressed with any one approach. Given that the staff's fundamental charge is to introduce and instill an appreciation for science in its constituents, there is an outward tendency towards a

positivist attitude about information. Staff teach students, and crave the ability to assert, truths about the world around us that are unequivocal. However, the equally powerful charge to show students their important place in this scientific world often hinges on telling stories about scientists from history whose truths have been shown to be inaccurate or flawed, even if still compelling and important; in other words to demonstrate the humanist side of scientific endeavors. With this objective comes the implication that even the truths science teaches today must always be in question, that these students themselves are the ones who should question those truths and unearth their own understandings, and that perhaps in their new answers to these questions, they will provide new truths for the future to question. And thus, it would appear that many of the educators, through the development of these programs, have become unwitting social constructivists (Vygotsky 1997), preaching the assembly of one's own understanding of the universe based not only on one's own experiences, but on the pushback to those experiences one receives from pitting them against someone else's.

These underlying themes in their programs drive Bakken staff to be naturally both critical and reflective. However, in recent years, evaluation has not been a welcomed or comfortable suggestion amongst its ranks. Within this era of Bakken staffing, few evaluations have taken place. Those that did occur were either large and unwieldy, producing little data of use to the people who had expended large amounts of energy on them, or were entirely unstructured and yielded skepticism as to the use or purpose of the practice. A select few staff members who claim more longevity at the institution than most still contend with memories of evaluations that produced anxiety and fear of individual recrimination in the successes or failures of programs. In short, staff did not like it and did not want it, but were aware that they needed it and thus were grudgingly considering it. The success of any given evaluation in such an environment is

dubious; for the practice of evaluation to take hold in an extended capacity approaches the impossible. Thus the challenge was not only to provide The Bakken staff with the ability to evaluate their own programs, but also to endow them with a sense of interest and investment in the process of doing so.

Posited Solution

A fractal is an image or graphic that, as one zooms in and looks on the micro level, it appears much the same as it does on the macro. A pattern, or structure, is repeatedly combined to yield a larger version of itself and this larger image is in turn combined as a component to form another, congruent, larger image in a feature known as self-similarity. No matter the degree to which one focuses in or steps back, no fundamentally new structure is revealed, only new re-workings of the existing structure (Mandelbrot 1983).

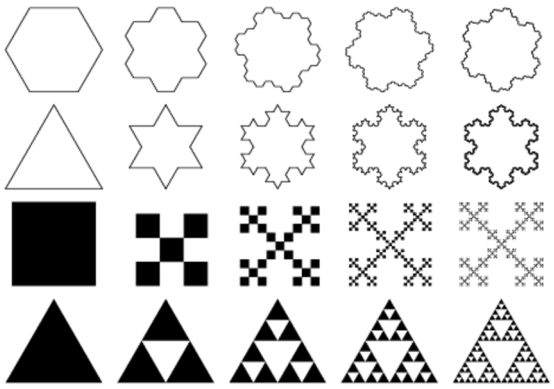


Fig 1.1 A series of fractal images (Wolfram Mathworld. 2015)

The notion of a fractal organization is not an original one. There is a current trend in business development to model organizational charts on this design. On their website, StrataGems, a consultant organization concerned with systems thinking in business, touts the virtues of a fractal system:

> This modern system enables organic processes of continuous improvement, innovation, and iterative development by changing member's perceptions about the relationship to one another and emphasizing constant communication (StrataGems, 2013).

While this description is not specifically addressing evaluation, the concept of practices that are organic, iterative, and improvement-driven does align well with existing approaches in that field, such as Michael Quinn Patton's Developmental Evaluation, described as a "learn-by-doing process" originated as an answer to a client's accusation that the categories of formative and summative evaluation were insufficient (2011). Educational programs are particularly well-suited to a design of this variety as, in general, their frameworks are already concerned with not only encouraging learning, but developing processes of learning as well.

In the field of education, and by extension the evaluations designed to assess education programs, such an arrangement could provide the necessary cultural relevance for program staff to accept and assimilate the practice. If an evaluation, or even an evaluation method, could be designed in such a way as to mimic or use the existing practices of a program, such pedagogical mimesis could allow that evaluation to walk the walk and talk the talk of a program itself. By reflecting the educational culture of the organization or institution in this way, an evaluation could produce results that are more easily understood, accepted, and used. Furthermore, if such a system could be designed and integrated or assimilated into the program, or even the organization as a whole, it could allow evaluation to become part of the cultural norm, to be seen as a self-similar part of its own fractal pattern. This re-framing would not only increase the program's capacity to assess and improve itself, but to cooperate and assist when the need for more comprehensive or advanced evaluations done by external sources might arise. Rather than super-imposing a structure upon a program's theory of action, a

fractal evaluation would build itself in the likeness of a program theory and become itself integral to the existing structure.

A fractal design system could offer a strong template as an evaluative structure for any educational organization, however, organizations concerned with science education would be particularly well suited to the format as they place an inherent value in seeing learning as process-based and, in most cases, have an articulated format for this process. This system, drawing upon the organization's own educational pedagogy to promote learning about the success of a program itself, would look to ameliorate several of the potential struggles of evaluation as posed decades ago (Weiler & Sterns, 1980). This model seems well-poised to provide a system that is culturally accepted and thus useful and used by the staff for The Bakken in particular, though this concept could likely easily be applied to other organizations and programs with similar missions. Informal science education is a growing community of those interested in teaching both concepts and comfort with science outside of a standard school curriculum (National Research Council 2009). This approach to having program staff conduct evaluations modeled after the processes they are already teaching students, could have implications in many of these informal science education programs.

Purpose of the Study

The purpose of this study was to explore the introduction of a fractal evaluation system, or a method that mirrored the existing teaching pedagogy, of The Bakken Museum's education department. The dominant intention of a fractal evaluation design is to increase comfort with, interest in, and ability to conduct evaluations. The study sought to understand how that implementation was able to affect staff conceptions and attitudes towards evaluation and effect change in the department's evaluation culture. The study, while done academically, also sought to provide useful information for continued

development of this concept, both inside The Bakken Museum and its education department as well as in the larger community of informal science education. This study can be summarized by one main research question and several sub-questions that contribute to answering it:

To what extent did the introduction of this evaluation system improve evaluation capacity in the education department of The Bakken Museum?

a) To what extent have attitudes and mindsets around evaluation in the education department changed?

b) What elements of this evaluation method potentially contributed to those changes?

c) In what ways have these changes to attitudes and mindsets affected evaluation practice?

d) To what degree are these changes likely to be ongoing and long-lasting?

CHAPTER 2: WONDER

What Do We Already Know?

Evaluation Capacity Building Literature

A study on instilling evaluation practice in one specific organization is not a particularly revolutionary endeavor. The desire to endow programs with not only the ability, but the desire to conduct their own evaluations or, barring that, to at least consider their work through an evaluative lens is longstanding and pervasive. While, historically, the job of evaluation often found its home as an externally motivated and conducted process, more recent attitudes have uncovered the problems with relegating the responsibility of asking and answering questions solely to outsiders. As Preskill and Torres state,

> Even when questions are asked in organizations today, there is a strong tendency to believe that expert knowledge resides somewhere other than within the organization itself. Although this tendency has created vast opportunities for external consultants who make a living by providing advice to organizations, it has had serious consequences for those within organizations who possess expertise that goes untapped (Preskill & Torres, 1999, p. 63).

There is typically an abundance of knowledge about programs and their minutiae existing amongst those who deal most closely with them and their constituents (Preskill & Torres,1999). The skills and desire to put that knowledge to work, to ask meaningful and useful questions, to collect and analyze information to answer those questions, and to provide an explanation or demonstration of the results of that process are what is often lacking (Pattison, Cohn, & Kollmann, 2013).

The interest in integrating program staff into the evaluation process is not a new one. Utilization-Focused Evaluation (Patton, 2008), Developmental Evaluation (Patton,

2011), and Practical-Participatory Evaluation (King, 1998) are just a few examples of approaches that have already recognized the benefit to using program staff knowledge and participation in evaluation. Alongside these traditions grew the field of Evaluation Capacity Building or ECB which concerns itself with instilling skills and structures to promote the practice of evaluation within an institution. There are many existing models that provide an approach to the introduction of ECB and its development. A particularly well aligned model, the Collaborative Immersion Approach to ECB (Huffman, Thomas, & Lawrenz. 2008), stems from a social constructivist theory of learning in which one gains understanding or makes meaning through the process of reflecting on experience (Vygotsky, 1997). This approach also asserts that this type of learning in an organization is best done socially, developing this understanding collectively, or collaboratively as the name would suggest, in teams of individuals with both stake in and understanding of the context at hand (Huffman, Thomas, & Lawrenz 2008).

Boris Volkov and Jean King also developed a simple and straightforward checklist for the process of ECB, as something of a synthesis of other work on the topic, designed to provide a framework of processes and considerations for introducing evaluation into an organization (Volkov & King, 2007). The checklist, derived from both literature review and case study, is an outline to guide considerations used to increase the long-term status of evaluation within an organization. The checklist includes points such as:

- ✓ Cultivate a positive, ECB-friendly internal organizational context.
- ✓ Build and reinforce infrastructure to support specific components of the evaluation process and communication systems.
- ✓ Build and expand peer learning structures. (Volkov & King, 2007)

This first point, the suggestion that attitudes toward evaluation must be positive if not welcoming to succeed in building capacity, and the second, that the development of structure that supports those attitudes through practice, provide a base to the argument

for this fractal evaluation model. As such, it should be stated that this model does not espouse its own techniques or methods so much as it provides a new framing of existing possibilities for such techniques and methods; in this case the evaluation would represent more of an appropriation of existing organizational structures to serve evaluation purposes, which could function as the infrastructure recommended by the second point above. The third point encourages fostering a community of learning, for which the process of evaluation could be central.

In further support of developing organizational context around evaluation, Preskill and Torres also assert that not only the process of conducting evaluation is productive, but so, too, is the process of sharing what is learned within the organization, enhancing the study to an example of organizational learning (Preskill & Torres, 1999). As this type of system becomes paradigmatic not just for programs and departments but for an organization as a whole, it even has the potential to turn its lens on the organization itself and all its itinerant operations and functions. To promote this concept more thoroughly, Preskill and Torres frame evaluation as a process of inquiry, a system of learning that can be used program- or organization-wide to increase understanding of programs' working elements and their successes and challenges. The practice of evaluative inquiry can allow an organization to fully integrate self-study into its identity and, through this, have evidence and support for the identity they choose to build and uphold. In a context where an organization can explore a range from the macro to the micro of its objectives using the same tools, the image of fractal would be self-evident.

Because ECB is prevalent and widespread in recent social science literature, there also exist a number of studies and articles attempting to summarize the findings and arguments of this literature. One such article is an extensive review of studies on bringing evaluative inquiry into organizational culture that suggests framing evaluation,

and thus ECB, as a system for inquiry and learning that becomes fully integrated into the structures, practices, and approaches of an organization (Cousins, Goh, & Clark, 2004). Through summarizing and combining the studies and practices of leading professionals in the field, they suggest that there are a number of dimensions, something of a progression of evaluation status, that outline the development of ECB in that context. Beginning with Evaluative Inquiry, an organization identifies a structure or process that allows staff to learn not only from evaluation, but about evaluation. Their next posited step is the development of those structures to instill Evaluation Consequences, or the continued practice of the processes of evaluation and the iterative and multi-faceted uses of evaluation activities and findings. From this state an organization can progress to Evaluation Capacity, wherein these practices and the understanding of them becomes endemic to organization structure, culture, and practice (Cousins, Goh, & Clark, 2004).

More recently, Labin et al. (2012) synthesized existing research and suggested, that, generally, while much of this research varied in language and focus, most of it indicated similar strands of thought. They developed the Integrative Model of ECB, which also sought to distill the literature into a single system, but is a model organized thematically rather than as the progression described above. This model suggests that the success of capacity building boils down to elucidating the Needs, the Activities, and Results for any organization, as well as how and where evaluation should play a role in telling the story of those categories (Labin et al, 2012).

Kristen Anderson wrote the most concise and straightforward summary of the literature on ECB that I have encountered in her Master's Thesis on the subject. Amongst the many themes she described as part of the reigning literature on ECB, there were several of notable relation to this project. Anderson discusses the need for the structures and activities that promote the organizational learning culture discussed

above (Anderson, 2013), but also references the necessity for positive staff attitudes about evaluation as discussed by Higa and Brandon (2008). Beyond this, Anderson also discusses the challenge to ECB of overcomplicated evaluation processes that are challenging for staff to understand, much less to complete (Anderson, 2013; Lennie, 2005). Furthermore, in her research findings on ECB in one particular organization Anderson discussed resistance to evaluation as being one of the major barriers to ECB. Anderson goes on to highlight thematic outcomes of Evaluation Capacity Building that include some more predictable staff outcomes, such as increased staff understanding of and ability to conduct evaluation, but also to use it to make decisions and improve programs (Anderson, 2013). These themes of staff attitude and acceptance of evaluation, conceptually, are the target of the posited fractal evaluation model.

Given that this study is also at an organization concerned with informal science education and that the suggestion at hand is that modeling their evaluation after said education systems, there is also some literature from that field that is of interest. The National Research Council suggests that informal science learning should be viewed as both a social and cultural practice, reminiscent of the Collaborative Immersion Approach to evaluation (Huffman, Thomas, & Lawrenz 2008; National Research Council, 2009). Beyond this the NRC proposes:

> ...an ecological framework for learning in places and pursuits intended to highlight the cognitive, social, and cultural learning processes and outcomes that are shaped by distinctive features of particular settings, learner motivation and backgrounds, and associated learning expectations. (National Research Council, 2009, p. 31)

While this framework was not aimed at discussing the learning of an organization or indeed the staff within it, it does speak to the same concerns around ECB that have been discussed above and is thoroughly aligned with the idea of develop a learning, or evaluation, process unique and organic to a specific environment and its learners.

Summary of the Literature

As ECB has been of intent focus and exploration in the field over the last decade, there has been a great deal written about it. Furthermore, there has been a fair amount written to synthesize that writing. An examination of these collections and summaries of the reigning theories yields many thematic elements to consider in Evaluation Capacity Building. Several of the recurring themes center on fostering positive and productive mindsets towards the practice in the organization. These themes provide support for this study and its further exploration of a possible manner in which to foster these mindsets.

Origins of Design

Fortunately, a fractal evaluation design for The Bakken, or realistically many other organizations with the same underlying mission and approach, does not require assembly from scratch. There already exist a number of approaches to evaluation centered around organizational learning or inquiry. One of these approaches would be apposite for this component and participatory method in likeness to the museum's own pedagogical approach. The Nanoscale Informal Science Education Network describes itself as a "national community of researchers and informal science educators dedicated to fostering public awareness, engagement, and understanding of nanoscale science, engineering, and technology" (NISENet, 2013). This collaborative network, largely dominated by museum education professionals, has designed and disseminated a large number of materials, programs, and exhibits to meet its mission. The Bakken, as a member of the network, has been a recipient of both funding and materials over the last five years.

Each NISENet program element underwent significant piloting and testing before being sent to over 200 member organizations. One of the unique opportunities of a

network of this breadth is the shared knowledge that arises from implementing a program element in a wide range of environments and locations. NISENet became aware that not all participating organizations had the capacity to conduct evaluations of these programs in their own locations, but the desire to understand the successes and challenges of the shared approaches was universal. Thus NISENet has developed an evaluation system, a system with roots heavily tied to longstanding evaluation methods such as action research, practitioner inquiry, learning-focused and participatory evaluation, and ECB (Pattison, Cohn, & Kollman, 2013). Team-Based Inquiry (TBI) encourages museum programs to derive actionable, timely, and self-generated evaluations through a process outlined as Question, Investigate, Reflect, and Improve (Pattison, Cohn, & Kollmann, 2013).

The Case for Team-Based Inquiry as a Fractal Evaluation for The Bakken

It does not take much scrutiny to observe the parallels between this system and The Bakken's Wonder, Try, Discover, Share (WTDS) method of student inquiry. The overlap between the two systems is quite apparent, requiring only a small idiomatic shift to completely align them.

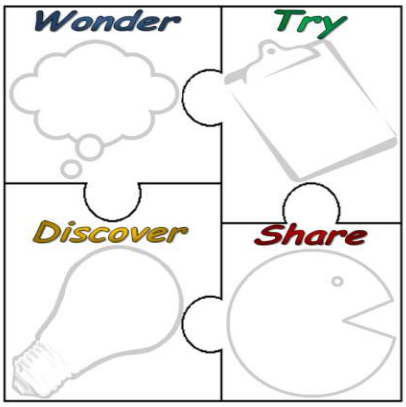


Fig 2.1
The Bakken Museum WTDS Graphic (Murphy, 2012)

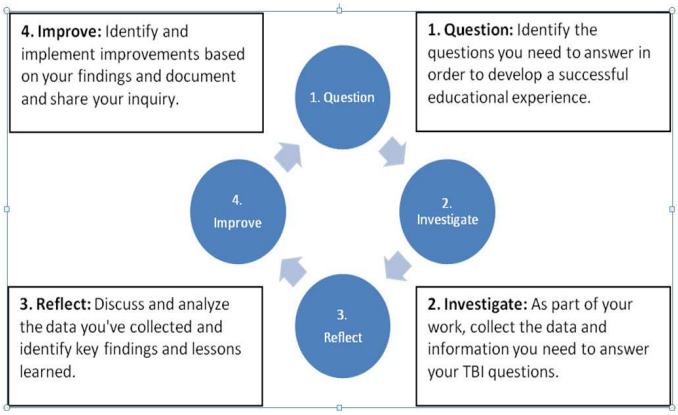


Fig 2.2
Team-Based Inquiry Graphic (Pattison, Cohn, & Kollmann, 2013)

Any informal science education program, and The Bakken's specifically, is already based in the promotion of inquisitiveness and curiosity. When considering a formative, developmental, or improvement-focused practice such as TBI, this curiosity can fuel the momentum of ECB as long the staff are able to see that the process is a medium to satisfy their self-derived questions and they are endowed with the tools to engage these inclinations in the service of evaluation. By making these tools feel less foreign, indeed self-similar or integral to what they do, this design could allow evaluation in the museum to be not only accepted, but to proliferate. If successful, this design would encourage Bakken staff to do exactly what they are encouraging their students to do.

Echoing Donald Campbell's call for an experimenting society (1984), for The Bakken Museum, a fractal evaluation could be the call to engage in its own organizational experimenting society of sorts. WTDS is already integral to the way the staff talk about the program objectives for learning science. By expanding the meaning and elements of each of these steps, program staff are well positioned to launch their own evaluative studies; the Bakken could not only learn to evaluate its programs, but also learn about evaluating its programs, developing a framework to integrate evaluative practices and mindsets for the long-term. These are elements that have already been shown to be central to ECB (Volkov & King, 2008; Preskill & Torres, 1999; Cousins, Goh, & Clark, 2004).

Determining the questions that are most pressing and whose answers will provide the most guidance about succeeding actions, or perhaps even succeeding questions, is a fundamental skill to evaluation (Patton, 2008). This notion is also similar to a concept that many Bakken programs teach. WTDS encourages students to ask questions that they, themselves, can and want to find answers to by experimenting; these are what they call exciting and investigable questions. Similarly, TBI encourages

staff to ask questions that are investigable, have unknown answers, and that provide timely, productive, and actionable information. As suggested by Michael Quinn Patton (2008), the skill set required to articulate these questions is one that often must be taught and requires informed reflection on the program, mapping out and agreeing upon the linkages between ideas, actions, and desired outcomes. By engaging in a process much akin to Patton's Utilization-Focused Evaluation, the ability to hone in on key elements of a program can reduce the unit of study to a smaller component and allows for asking and answering the truly important questions in a more manageable manner (Patton, 2008). This type of specific and focused examination can often result in more apparent and actionable answers. TBI maintains very similar premises, suggesting that each evaluation attempt should bite off only as much as it can reasonably be expected to chew, consider, and swallow in a timely manner given the resources and concerns of that program and its staff at that time.

Ironically, this application of an elementary scientific process to evaluating a science education program could cause advocates of scientific experiment-based quantitative evaluation some concern. Although evaluators or researchers experienced with more qualitative or participatory practices would likely accept TBI more easily, none would argue that that this type of study would have any sort of robust external validity. Internal evaluation, in and of itself, seems to have little external validity as few people or policy makers will take at face-value a program staff stating that they have evaluated their own program and shown it to be wildly successful. However, the aim of this method is entirely focused on specific primary intended users, i.e. educators and designers. They are not primarily concerned with trying to prove to others their worth, but rather with knowiing that they themselves are providing the best product they can in service of their mission.

The desire for information about a program's successes and challenges is both internally and externally motivated. It is true that, in many cases, funders require evidence that the resources they provided were used appropriately and effectively and in some cases a more formally conducted evaluation than I am suggesting may be necessary (Pattison, Cohn, & Kollmann, 2013). Nevertheless, the ability of an organization to demonstrate that their programming has been developed and conducted using data-based decision-making processes would likely strengthen its appeal when seeking assistance from those funders.

This design is meant to be a function of the organization itself, and while it is one that could be shared externally for leverage, it would fundamentally be in place to ensure that programs are doing the best they can as they are developing or happening. It is this underlying intention that may ease potential concerns regarding validity. There is no reason for a program staff, endeavoring to achieve some educational objective, to falsely inflate their own abilities to themselves. Indeed internal evaluation, being an extension of self-reflection, may in many cases yield a more honest review of successes and challenges. As King (1998) states, "American pragmatists' approaches stem by nature from social meliorism, and to ignore their meliorism is to misrepresent their intentions." (p. 65) Even proponents of the theoretical apex of quantitative methodological exactitude, the randomized control trial, support the notion that those who have a vested interest in a project are perhaps the best to study the meaning of its value or interest. As Donald Campbell (1984) intones:

> The objectivity of physical science does not come from turning over the running of experiments to people who could not care less about the outcome, nor from having a separate staff to read the meters. (p. 35)

The investment and understanding of program staff can serve to increase the insights delivered by an evaluation. One of the benefits to an internally run program

design is its flexibility. Team-Based Inquiry emphasizes the large variety of useful information and data that can be gathered, as well as a plethora of employable methods, to inform decisions about programs.

Within the context of this fractal evaluation, many small program components could be explored. Smaller program elements can be viewed with a smaller scope, are simpler to assess, and are more nimble in that they can more easily be altered or experimented with than entire programs. The actions taken after the discovery phase of a TBI study like this can range greatly including the decision to discontinue a program component or a practice within a program, selecting one of a series of planned-variation options as “what works better” (Boruch, 1991), making minor formative changes, or even identifying additional questions. Each of these actions may, in turn, also result in an iterative re-evaluation of changes to determine the success of the action taken. The prescriptively short time frame and relative simplicity of research methods allow the evaluation to avoid being an unwieldy or one-shot endeavor, rather to be several steps in an ongoing cycle of program-based improvement instead. The results of these methods could be viewed as being usefully semi-formal, as described by the architects of TBI, Pattison, Cohn, and Kollmann (2013) as “part of a spectrum, somewhere between mucking about and formal evaluation” (p. 6).

It is the system of analysis for TBI that is perhaps the feature that most effectively positions it for adoption by The Bakken education staff, supplanting their “Discover” step with TBI's “Reflect.” Team members for any given evaluation sit down and examine the data en masse, first identifying data points that are most striking to each participant, then delving into parsing and grouping patterns and connections as they are observed. This is a structured process that allows participants to collaboratively construct meaning from the gathered data, as is suggested by previously discussed social constructivist theories

of ECB (Huffman, Thomas, & Lawrenz, 2008). Team-Based Inquiry, and the traditions of UFE, P-PE, and action research from which it was derived, would encourage staff not only to carefully articulate their program's processes and objectives, but also examine themselves and make their own meaning from the findings (Patton, 2008; King, 1998). Thus staff would be able to not only answer whatever evaluative question has been posed, but to determine what questions will be posed next; the iterative nature of the process or cycle serves to enhance ongoing organizational learning as a whole. It is this social learning, this curiosity, this experimenting community that The Bakken strives to create in its mission; the organization would be more familiar with, more interested in, and more amenable to such a communal and social method of evaluation and analysis, as it would challenge the staff to practice the very method they preach.

Summary of the Case

Team-Based Inquiry provides a strong model that is already configured in such a way as to make it align effectively with the program approach and objectives of The Bakken Museum's education department. This alignment, as well as the strength of the model itself, will help the implementation of evaluation seem more comfortable and organic to program staff and thus could provide not only the tools and skills necessary to conduct their own evaluations, but also the perceptions and attitudes towards the process that will make the endeavor become an intrinsically motivated one that will truly build an organizational capacity and culture around evaluation. These tools, skills, and attitudes, much like the Science Assets identified by The Bakken, are also imperative to the ongoing success of evaluation in an institution. These elements, or the Evaluation Assets as I will refer to them now, are listed below.

Evaluation Assets	
Tools of evaluation	Model or method
	Resources
	Support
Skills of evaluation	Identifying need for evaluation
	Determining evaluation questions
	Designing data collection instruments
	Collecting data
	Analyzing or interpreting evaluation data
	Making or recommending changes based on evaluation data
Attitudes regarding evaluation	Evaluation is doable
	Evaluation is useful
	Evaluation is worthwhile

Table 1. The evaluation assets

The development of the first two assets listed in the table has been addressed in myriad ways through various different approaches and systems for ECB. This study, while examining those assets to some degree, is primarily concerned with the third asset listed, attitudes regarding evaluation, and how this fractal model could enhance those attitudes while also building the tools and skills.

CHAPTER 3--TRY

What Did We Do?

Implementation of the Method

In December 2012, I attended a National Meeting of the Nanoscale Informal Science Education Network (NISENet) in Boston. While at this conference, I attended a session in which the presenters introduced the concept of Team-Based Inquiry as a simple, timely, and action-oriented evaluation system. As previously stated, the similarities between this system's design and The Bakken's own version of the scientific process were striking. As the museum, and specifically the education department, had no formal, consistent, or standard approach to evaluation, it seemed an excellent fit. I was already involved in coursework for a Masters in Evaluation Studies and felt that I had enough of a base of information to help introduce and guide this process for the museum, and what information I did not have, I had ample access to through further coursework and contact with professors and practitioners. Furthermore, my suspicion was that the similarities of the system to the education department's teaching pedagogy might allow this practice to take hold with the staff, to become a standard and self-driven practice rather than to be a burr in their side or an onerous, required task.

There was also a particular opportunity at hand, as there was a Bakken education program for which funding had been reduced, and the program needed to cut from 3 classroom visits to only 2 visits. Working with the staff of this program, I introduced the concept of TBI, and thus began the first study, determining which program elements seemed to have the greatest impact and thus should continue to be included in the shorter program. The study yielded productive results that were understood, accepted, and acted upon. After the success of this first TBI study, staff in

other areas of the education department became interested in the process, and TBI was introduced to them as well. While there was some necessary skill-building in this introduction that took time, that time was budgeted with enthusiasm; there was very little resistance to the process or to the system. Many staff stated that it seemed obvious, but still useful. Less than 3 years later, more than 8 TBI studies have been done in the education department and there are more in planning.

Examples of TBI Studies done by The Bakken Museum Education Department

Science Assets Outreach Program

The Science Assets Outreach Program was a three-day residency program with 4th grade students in the Minneapolis public schools. The program, while addressing some elements of content around electricity, was primarily focused on building student Science Assets, or attitudes and thinking skills. This program was not fee-for-service and was funded entirely by grant support and district contributions. In early 2013, it became clear that the funding structure for the program was changing, necessitating large budget cuts. As the program had experienced many successes, both in appreciation of students and teachers and in quantitative measures of its objectives, The Bakken chose to continue the partnership with Minneapolis, but sought to shorten the program to relieve the budgetary strain. Faced with the endeavor of cutting a 3-day program to two days, staff were struggling to identify the most useful elements with the highest impact for students. Team-Based Inquiry was an optimal tool to help program staff make that determination.

In the spring of 2013, as part of a separate but related evaluation taking place, Bakken staff returned to classrooms that had experienced the program the previous fall.

Staff provided students with a blank piece of paper, and asked that students write down everything they remembered about The Bakken residency. Responses were gathered and re-typed, then examined by staff and discussed and coded for what activities they pertained to, as well as any qualifiers that students may have chosen to attach to their description. The codes were graphed to examine prevalence of themes, but responses were also considered in their entirety by the team, as is indicated by the TBI Reflect process (Pattison, Cohn, & Kollmann, 2013).

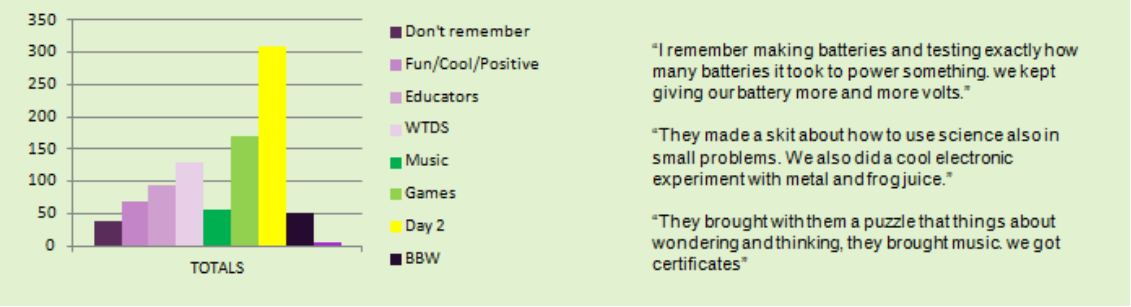


Fig 3.1 Student recall of program activities (Bakken 2013)

The data provide a distinct picture of what program elements had the greatest long-term staying power for students, and also provided several other useful pieces of information for program design changes. For example, in the program above, Day 2 centered around an experiment about the engineering of the battery and a puppet show about the historical scientist who first performed that experiment. Both the experiment and the story continued to resonate with students after several months at a notably greater rate, being referenced by more than twice as many students as the next most recalled activity. The Day 3 elements were barely referenced at all and the Day 1 elements were somewhere in between. Upon observing this in the data, program staff elected to keep Day 2 completely intact and select just a few key elements from Day 3 to integrate into the Day 1 curriculum, effectively cutting Day 3 from the program.

Science Assets in Field Trip Programming

After the success of the Science Assets Outreach study, the onsite field trip program at The Bakken opted to use TBI to help determine the success of its recent implementation of the Science Assets as a major program objective. A team of four staff was formed and, in the interest of keeping the study very simple, a single question was identified that staff felt would be informative about student attitudes: "Do you think like a scientist?" The question was posed and the process described to the students on the bus upon their arrival to the museum, and each student was given a popsicle stick. As students walked through the doors they passed a cart that had three buckets labeled with Yes, No, and I Don't Know. Students placed their stick in the bucket which corresponded to their opinion of themselves. The process was repeated after their 3.5-hour program experience, as students exited the building. A later iteration of the program color-coded the sticks for gender, as well. The following data were generated:

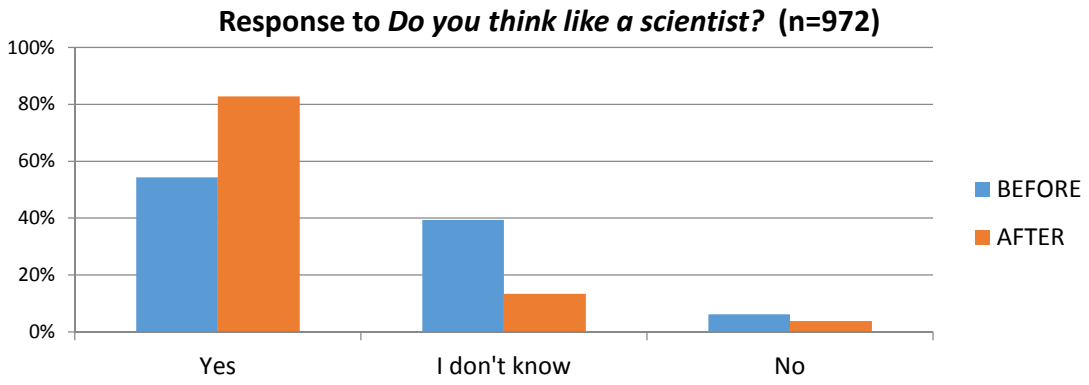


Fig 3.2 Field Trip Student Responses (Bakken 2013)

Formally, these data were examined and discussed as an entire staff, and used to determine what other questions staff were interested in exploring. Informally, many staff found themselves checking the numbers as they were tabulated at the end of a day's program when they had been teaching, and used them as an instant feedback metric to experiment with slight adjustments to their teaching approach. Beyond this, the data

were ultimately used in numerous grant reports and applications despite that not having been the original intent.

Outreach Nanoscience Assembly

In December 2013, the outreach program of The Bakken was awarded a grant by NISENet to develop an assembly show about the topic of nanoscale science. As an additional supported opportunity, two staff from the program were invited to participate in a Team-Based Inquiry Cohort with teams from other institutions across the country who had also received NISENet grants. The cohort participated in trainings on TBI as well as collective conversation and planning around how it would be used to evaluate and report on the success of their grant projects.

As a result of this, The Bakken Outreach Program sought to develop an evaluation of the new assembly show to determine increases in content understanding as a result of the program, based on a model developed by Holly Walter Kerby and Fusion Science Theater (now Fusion Science Learning). This model capitalizes on theater and storytelling structure, as well as recorded and matched pre- and post-predictions that both increase audience investment in the progress of the show and provide evaluation data about content learning (Fusion Science Learning, 2014). Using this model and the TBI training as a guide, the team built the show with this evaluation element embedded and scheduled five pilot performances. The program design involved setting up an experiment and, before demonstrating it, asking audience members to predict the outcome on sheets that had been numbered.

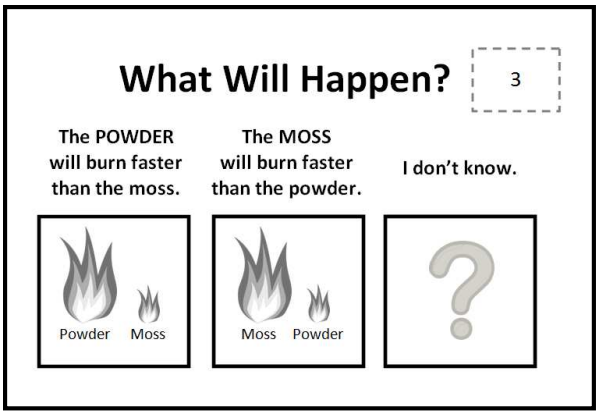


Fig. 3.3 Audience prediction sheet

The demonstration was then delayed through a series of comedic interactions between educators through which the scientific content around the experiment was delivered to the audience. After a series of related demonstrations and discussions, the audience was asked make their prediction again on a sheet with a matched number to their previous guess. Only after this did the originally posited experiment take place. Examining the two matched predictions allowed for something of a pre- and post-comparison of knowledge regarding the content taught in the program.

Upon examination of the data generated by the first show, staff immediately identified weaknesses and made improvements before the second show was performed. This process was repeated again after the fourth performance. The success of the improvements was visible in the data produced by each show:

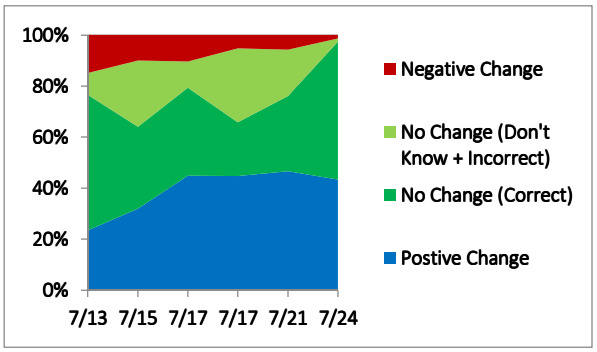


Fig 3.4 Change in audience understanding (2014)

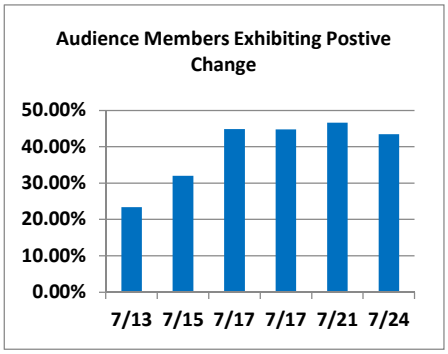


Fig 3.5 Percent showing improved understanding (2014)

Simple interviews were also conducted with adult audience members after the performance of the fifth show, asking what they liked about the show, how they thought the show could be improved, and to describe in their own words what the show was about in order to gain more detailed insight into the audience experience of the program Responses were examined and used to make further improvements. Many audience members responded with comments about the visibility of one the demonstrations and about the desire to understand the applications of the science discussed. The demonstration was re-created in a larger form and the script altered to include a discussion of applications in response to these data. An unanticipated result of the interview was that many audience members commented that the show was not only about nanoscience content, but also about the scientific process itself and the necessity of making guesses and observations as well as being able to change one's mind once more was learned. This information was then used to shift the program description in marketing materials and has since been a popular point of interest to schools and teachers. The outreach team compiled all the data and used it to create a report that was presented to the rest of the TBI Cohort in September 2014; the information was also included in the grant reporting for the program.

Ongoing Studies

The staff of The Bakken education department has continued to use TBI in a number of studies beyond these three. Several of these studies have been ongoing and iterated processes, wherein the study answers one question, but produces others, which are then examined. More studies are also in planning. As of the spring of 2015, these studies have been quartered to only the education department, but there has been discussion of expanding and disseminating the model to other departments in the

museum. Two new exhibit installations over the next two years could offer prime opportunities for this method to find use beyond the education department.

Reach of the Project

The intention of this study was to understand to what degree the introduction of TBI, as system which so thoroughly mirrored the pedagogy of The Bakken Museum, improved the organization's, or specifically the education department's, capacity to do evaluation. Early in the study, however, the understanding of what elements of evaluation capacity were within the purview of this implementation and thus of interest shifted somewhat. While the implementation did involve some development of skills for staff, the degree and depth of knowledge remained specific to the smaller scale intentions espoused by Team-Based Inquiry. A thorough evaluation capacity building intervention would require, as already discussed, a more comprehensive training and more complex structural integration than was within reach.

The Science Assets that are taught at The Bakken are not only about technical ability, but also about fostering the willingness to engage with the field and a belief in its value. Similarly, the Evaluation Assets are these same elements including tools and skills, but also the development of staff belief in their own capability to conduct evaluation and understanding of the worth of the endeavor. In this case TBI provided most of the needed tools. This study was focused on uncovering was how the discovery and practice of TBI impacted the attitudes and mindsets around staff ideas and approaches to working evaluatively. Evaluation as an internalized, endemic organizational behavior will not survive without interest in it and the belief that it is useful, practical, feasible, and necessary.

CHAPTER 4--DISCOVER

How Did We Know?

Framework for the Study

The impetus behind conducting this study using this particular evaluation model, in this particular department, of this particular organization was dual. As previously described, this method and department had strong parallels that required very little effort to draw, and The Bakken's Education programs were in need of evaluation. Beyond this, as a Master's Degree Candidate in Evaluation Studies at the University of Minnesota, an active member of the NISENet, and a full-time employee of the education department of the museum, I was effectively positioned to have the skills, knowledge, access, and impetus to implement and examine the results of Team-Based Inquiry at The Bakken. The study was begun, not as a study but as an employee-driven endeavor to improve data collection and decision making at the organization and in the education department. Early on, however, as the introduction of TBI began, I became aware of how problematic the culture around evaluation had been. I observed that the elements of the Evaluation Assets, as they would become, were not much in evidence. The opportunity to use this model, one that was, essentially, the same one that Bakken staff used with students, to examine a possible approach to improving those assets caused this project to evolve from an organizational endeavor to one of research.

The scope of this study was small, comprising only one organization and, for the most part, only one department of that organization, and focusing on only one topic within that scope. Because of this there were elements of a case study in the process, although, as a more thorough exploration of other components of the case w was not of interest, it does not fully meet the charge of a case study (Creswell, 2013). Given the

concern with an adjustment in focus to the attitudes and mindsets of staff, there is something of a phenomenographic approach in use. Phenomenography describes a methodology that attempts to define the breadth and nature of perceptions that a given group of people have about a particular phenomenon (Marton, 1987). If, in this case, we allow the idea of evaluation to be the phenomenon in question, this study is similarly trying to glean the attitudes about that phenomenon and how those attitudes function in the world of practice. The extent and effectiveness of the department's evaluation and TBI activities were a matter of record. The needed information was how TBI had affected programs and program staff in the ways are not discernible from documents and records, but also the ways that were perhaps most crucial to the survival of the practice—or the attitudes and mindsets towards the practice comprised by the Evaluation Assets. Thus the charge of this study was to uncover these attitudes and personal assessments about evaluation, about TBI, and about how TBI may have contributed to any changes in staff perspective.

Design

To explore the research questions of interest, the following study design was created:

Research Question	Information Needed	Data Collection Method
To what extent did the introduction of this evaluation system improve evaluation capacity in the education department?	Record of activities, Insights of staff	Document Review, Interviews, Survey
To what extent have attitudes and mindsets around evaluation in the education department changed?	Insights of staff	Interviews
In what ways did the introduction of this particular evaluation method potentially contribute to those changes?	Insights of staff	Interviews
How did these changes in attitude and mindset affect evaluation practices?	Record of activities, Insights of staff	Document Review, Interviews, Survey
To what degree are these changes likely to be ongoing and long-lasting?	Insights of Staff	Interviews, Survey

Table 2. Research methods by research question

Ethics

This study was approved for an IRB Type 2 Exemption from the University of Minnesota Institutional Review Board on February 5th, 2015.

Data Collection

For all participants in this study, informed consent was obtained, verbally for interviews, and using an " Accept" button for the online survey. Names of participants were kept separate from their interviews, as well as names they may have referenced while speaking. There were no identity markers officially attached to the survey information, beyond people's participation or non-participation in TBI, although some respondents chose to make reference to identifying information in their answers. Those references are not directly included in the report of results.

Interviews

The primary method for data collection in this study was the semi-structured qualitative interview. A protocol for these interviews was established including main questions, likely probes, and one activity. This interview was designed with a responsive interviewing model, as described by Rubin and Rubin (2012), chosen for its flexibility and the opportunity for the interviewer to engage in pertinent avenues of questioning as they become evident during the course of the interview. This interview protocol was piloted in fall of 2014, as part of a class on qualitative research methods.

The interview process was centered on seven main questions (Appendix A). The interviewer asked probe and follow-up questions whenever it was warranted to elicit greater detail or depth. Some questions were designed for straightforward response,

while others were used to elicit opinions or emotional responses through description. As part of the interview process, participants were also asked to draw a picture expressing their perception of evaluation as a broad concept. Participants were encouraged to be as concrete or as abstract as they liked, with the only constraint being that it had to fit on one side of one 8.5 x11 piece of paper (Appendix B). After the interviews were conducted, they were fully transcribed and analyzed for themes which were coded, summarized, grouped, and regularly compared with each other to continuously ensure accuracy and coverage of the codes (Rubin & Rubin, 2012).

Interviewees were selected based on past participation with at least one TBI study in two out of three capacities: design, collection, or analysis. Given that the study was examining the effects of TBI implementation, it was necessary to interview respondents with significant enough involvement and understanding of the process to be able to comment on TBI and those who had the opportunity to experience effects of its introduction, be they perceived by the respondent or no. As there were only 11 individuals who met these criteria, they were all recruited to participate; 9 individuals agreed to an interview. Of those 9, 5 were female and 4 were male, and the group comprised multiple positions within the department, including 1 part-time educator, 3 program coordinators, 3 program managers, and 2 program directors. Years of employment at the museum ranged from 6 to 15.

Surveys

Additionally an electronic survey was sent to all current staff of The Bakken. The survey used mostly quantitative questions to collect basic information about staff experience levels and attitudes about evaluation (Appendix C). There were a few open-ended qualitative questions to allow staff to more thoroughly describe their perception of

evaluation and its potential use at the museum. A yes/no question about TBI participation was included to allow cross-referencing to those who were TBI participants versus those who were non-participants. The goal of the survey was to create a comparison between perceptions of individuals in those two groups, but also to provide information on the perspectives of the museum as a whole entity and possibly inform next steps in disseminating the TBI method or an approach to evaluation more generally, beyond the education department, were that to be indicated by the results of study.

There are only about 30 staff at The Bakken Museum, total, thus any sort of sampling would be impractical. The survey was sent to all staff members as a link via a recruitment email; 20 of those 30 participated. Given that 7 of the original 30 staff members are new employees to the museum working only 10 hours a week, this was approximately the response rate that was anticipated. As there was not identifying information attached to the survey, there is no associated demographic information for given responses, other than that 11 respondents were TBI participants and 9 were not. There was no direct repetition of questions, so data were not overtly repeated between surveys and interviews.

Document Review

Records and reports regarding the results of the TBI studies conducted by the education department were available through requests from staff and access to the company server. These documents were reviewed for evidence of the nature and effectiveness of evaluation practices using TBI. To protect information of a private or sensitive nature, information gleaned from the documents was only reported on explicitly where it had already been shared in a public capacity.

Limitations to Study

Positionality

While this might initially seem to present a conflict of interest, that does not take into account the actual motivation in this work. The interest was not to uphold The Bakken as a beacon of evaluation capacity brilliance, but to improve the form, function and feeling around evaluation in a particular department and, possibly, an organization. This study is meant to be a telling of this story in a way that could allow the experience of this organization doing this work to be of use in addressing challenges for other organizations beyond this one. In many ways this was a much more complex version of a TBI-like study (minus the team aspect) on the introduction of TBI to an organization. Much like in other TBI studies, this is not the end of my interest in this case. It is my underlying goal to see this process succeed, and thus there is equal need to uncover the challenges and weakness as to highlight the successes. It is my hope and intention with this study to provide not only reflective, but actionable information for the future of evaluative inquiry at The Bakken Museum.

I was very close to this project, outside of this research, and that poses several potential problems with the data collected. As the person who introduced TBI to the staff and the person conducting interviews about that process with the same staff, there is a possibility that participants would speak more positively of that particular concept out of respect or kindness to me. They may have been concerned that criticism of that process would offend or upset me. Furthermore, they may have chosen to speak more positively than they actually felt when giving the project praise. To try to diminish this concern, I chose to ask a number of questions in which they could describe processes and actions without specifically requesting qualitative assessments, as many of these descriptions elicited these qualitative opinions in the re-telling and in a manner that would not be

interpreted as criticism of me or of the system directly, but was more descriptive of their personal experiences.

Beyond this, all the studies and processes they spoke about had been debriefed as they happened, well in advance of the interviews. Each respondent had already had the opportunity to provide both positive and negative feedback on the process in an environment where I would not have been in a position with a notable power dynamic, but for which I was present, allowing me to roughly triangulate their interview responses to impressions of previous given commentary as a means of comparative cross-validation.

The advantage of my native status with my interviewees was that I did not encounter much reluctance in the process. My respondents appeared comfortable with me and very willing to talk easily with little prompting. Furthermore, we all share a lexicon; I did not need to expend time in clarifying various referents as I already had an understanding of those terms. This could, at times, prove challenging, as I had to occasionally coax respondents to speak specifically rather than using a gloss or assuming that I knew what something was like and would not need or want a description, but mostly allowed interviews to take on a more comfortable, conversational tone.

There still exists the possibility of bias from my respondents. There also exists an obvious bias in my own interpretations. The study must include a bracketing of my own preference to see the project succeed.

Scope

The scope of this study was very small covering only one department of only one institution. Furthermore, the time-frame for data collection was very short—4 weeks—

and thus did not allow for conducting multiple interviews, encouraging more complete participation from non-respondents, or more in-depth exploration.

Participation

The pool of participants for this study was also very small. While all but two staff members who were eligible for interviews chose to participate, survey participation was somewhat weaker. A possible result of this is that respondents who chose to participate could likely be those who already had strong attitudes or beliefs about evaluation, and thus it may not capture the perspective of staff who feel apathetic or uninterested in the topic. This lack could not only skew the data, but also misses an opportunity to examine how those attitudes and mindsets towards evaluation could be productively shifted in a positive direction.

CHAPTER 5--SHARE

What Did We Find?

Qualitative Interview and Survey Results

The examination of the interview transcripts revealed a number of prevalent themes that inform answers to the research questions. Open-ended survey responses were all included in this analysis. When those survey responses were identified as being participants in Team-Based Inquiry they were included in the review of interview information. Non-participants' responses were considered separately. This chapter begins with a discussion of themes from TBI participants.

Themes from the Pre-TBI Era

Interviewees were asked to reflectively describe evaluations that were happening at The Bakken prior to the introduction of Team-Based Inquiry, as well as their relationship to or roles in the evaluations of which they chose to speak. From these descriptions and reflections, a number of themes regarding what evaluation was and how people felt about the practice during that time became evident. A number of respondents initially claimed that there had been no or little evaluation happening. However, in each of those cases, they ultimately amended that assertion to describe some practices and attitudes that had been in place, even where those practices did not yield, in their opinion, much information of use.

> So we accepted everything as it was. There was no movement toward change. 'Though I would say there was curiosity. We are all people who appreciate science and we do all like to know how things work. But it had never really been approached any other way, so we left it at that.

The following are themes regarding the experience and perception of evaluation for participating program staff.

Evaluation was perceived as either very formal or very informal

Most descriptions of the evaluations that had taken place at The Bakken prior to the arrival of TBI were described as being either very formal or very informal. There was little recall or sense of evaluation happening during that time that was not relegated to one of those two extremes.

Formal Evaluation

These evaluations were spoken of as being strict, stuffy, and typically externally-driven, if not entirely motivated for the purpose of appeasing funders or writing grant reports. These evaluations often seemed distant from actual program activities due to this formality.

Formal evaluation was done for and/or by external sources

Formal evaluation was perceived as being mostly orchestrated by outside sources, involving strict rules that were set down for program staff, who were called on to participate in administering and grading the instruments, although often not in planning for or analyzing the results of the studies. One staff member acknowledged that there seemed to be a belief that this type of external control was necessary for validity.

> And I do think that although it shifted, I am sure there is still a remnant of the old style of evaluation where it is believed that you need to have the objective observer behind the one-way mirror. That idea that that is the only way you are going to know if your program has impact is if a person who is looking at the program has nothing to do with the program. Which doesn't seem, from the evaluation field, to be a preferred method anymore. But that's something that many of us have experienced.

There was also an expressed perception that this type of evaluation was only done to appease funders or grant panels and was not particularly aimed at use within a program.

Formal evaluation was unpleasant to conduct

Many respondents referred to experiences with large scale, quantitative evaluations they had participated in and described their experiences in data collection and analysis with descriptors such as tedious, agonizing, awkward, and time-consuming. The rigidity of these evaluations and the processes surrounding them was often cited.

> I mean this survey was super formal, because I think it was for a grant? So we had to do it…this certain way. So, yeah. Scripted....the process was, well it was agonizing.

Formal evaluation had debatable validity

Staff who had worked on larger and more formal evaluation efforts expressed discomfort with quantitative instruments' ability to generate accurate, complete, or deep enough information to truly understand the complexities of their programs or of student learning within those programs.

> Yeah, well…we were distilling such complex…I mean kids learning and their thoughts have so many different sides and elements and we were smashing all of that into a number, or into a few numbers, but they couldn't actually show what the kids understood or thought, at least not completely. And half the time we weren't all that sure of the number anyway. So yeah, it was the part where we were fitting something big and messy into something small and specific.

Informal Evaluation

Other forms of evaluation were done in a manner that many described as being so informal or so casual that they did not perceive it as evaluation at the time. Staff

indicated that informal reflection was something of a standard practice, but there was no standardization to it. While these practices were described as useful most of the time, they also tended to have descriptors such a fuzzy, casual, or sloppy.

Reflective practice has long been common amongst education staff

Several people spoke of the education staff's long-standing ability and willingness to think and talk about their programs. This was described as typically being very casual and largely undocumented. This mindset was something people felt was an asset to the department.

> And in part because my teaching experience had been in a public school setting, so to me it seemed highly reflective, highly best practice. Because in the settings I had been in it was rather individual. So I was sort of, I really lauded The Bakken for having a culture of self-reflection and awareness even though it wasn't done too formally, or even in some cases as effectively as it could have been.

Staff perceptions on the necessity of reflection for program growth were strong, but typically were not associated with any particular practice or process to acquire or engage those reflections.

Evaluation was mostly perfunctory

Most staff indicated that they believed and understood evaluation to have been necessary to acquire and appease funders and as a gesture of "best practice."

> To be honest, it seemed more like we were doing most of the surveying because we felt like it was… "best practice" to do it, but we weren't looking too carefully at how to do it. Just having it done was most of what we were going for, in the end. Which is I guess, not that useful.

Many held the notion that evaluation was something they were supposed to do so that it could be said that it had been done—so they could "check a box"—and that its use typically did not extend much beyond that.

Evaluations were repetitive and platitudinous rather than insightful

There was a described pattern of evaluation asking the same or similar questions repetitively and that many of these questions were not designed for decision-making or improvement, but rather to provide a “pat on the back.” This often meant that data collected for these evaluations was not given much bearing or value after they were acquired.

> Well, it got to the point where we kept asking the same kinds of questions so we weren’t really getting anything out of it. It was more just reaffirming what we already knew. And it didn’t really provide us any new insight.

Evaluation data disappeared after they were collected and were never used

In many cases, respondents struggled to recall or describe what had happened to evaluation data after it was collected. Many respondents talked of data vanishing upon collection, with repeated descriptions of them being “put in a drawer” or “collecting dust.”

> I think it's too easy to generate the data, and even to do the math, do the calculations and then shrug your shoulders and say that's enough, or pulling more from that is too much, let's put it a drawer.

There were references to the "number-crunching" or data analysis happening somewhere else or by someone else, but results were not often seen again by program staff. This contributed to a perspective that evaluation was time-consuming and not particularly productive.

> There is also maybe a perception of overuse of evaluation, or of sort of meaningless evaluation. Analysis for analysis sake—or analysis paralysis.(Laughs) More than is necessary and the sort where the data that just disappears [sic] after its [sic] collected. It just gets done to say it was done.

Evaluation data were not produced until it was too late

Respondents also suggested that evaluation almost always took on a summative nature, with information only becoming apparent well after a program's completion and often after it had been discontinued or already changed.

> But at any rate we often weren't getting the data results back for months and months. So then we went, "Oh! Isn't it interesting that we just did that at every single school and we didn't know that this thing was happening that we might have chosen to change it had we known earlier." In some cases it was like "Okay we'll change it for next year." And I think it at least in one it was like well we would change it next year except the program itself has to cut a day, or whatever happened. So it was never able to be used.

In the eyes of program staff, this perspective decreased both the evaluation's value and, in some cases, the morale of staff who discovered that they had not achieved their objectives as effectively as they had hoped at a point in time when there was nothing they could do to correct it.

> Well, I mean, as an educator you wouldn't want to know after the fact that something you were doing wasn't…landing with students. I would so rather know in the middle…or at the beginning! (Laughs) But then you could have a chance to play with it, to try something that works better.

Themes from Post TBI Era

Interview participants were also asked to describe and discuss evaluations that had taken place more recently using TBI as a method. Once again, respondents were asked to describe the process and goals of the evaluation as well as their role and experience with that study. Themes produced reflect the staff perspectives and experience of evaluation since this change in practice.

Evaluation is useful

Staff discussed evaluations that had happened since with notably positive descriptors, generally, and in contrast to their previous experiences with evaluation specifically. They

also outlined a number of perspectives on evaluation that demonstrated a shift or change in the perceived status of evaluation in their minds.

Evaluation provides data to make decisions/improvements

Every respondent referred to a least one decision or program improvement that had been made using TBI and how the TBI data assisted in that process, both when considered by program staff and when needing to make a case for changes to constituents outside the department such as justification for not raising program prices or advocating for a particular staffing structure.

> And when we are trying to makes decisions or changes, we all had the reasons to back up what we are saying. Sometimes arguments we have been making for a while. Sometimes new things. But we seem to…know more now when we make a choice.

Evaluation can be systematic and simple

There was a prevalent discussion about a new-found understanding of the opportunity to systematically generate data in ways that were more simple and practical than those previously experienced.

> We needed quick and simple. So it was really cool that we came up with a system that was really pretty straightforward and really successful…reliable…easy to replicate. It was certainly beneficial.

Evaluation can be actionable

The concept of evaluation design to generate data that answer specific, pressing questions or direct concrete programming decisions or improvements was compelling to many staff.

> The fact that we are using the data. That we came up with our questions that we wanted to ask, that we were asking a question that was going to be something that had an effect that we could act on.

The consideration and construction of these questions was a logical, but mostly new concept for most respondents.

> And then also learning, one of the skills I had to learn in going through this process, is do you want to ask questions to which you can do nothing about? When you find out what the answer, is it something that we, in our positions, have any power to effect or make changes. If not, is it worth asking the question?

Evaluation was timely

There was extensive discussion of the reduced timeframe provided to staff by these new methods. The ability to see discernible results expediently made the process of collection seem more obviously productive.

> And then we counted everything up at the end of each program, so we saw what was happening right then and there. And we did actually work that into discussion for the staff.

Beyond the convenience of acquiring understandable data quickly, staff referenced the opportunity to be more experimental and varied with their teaching practices; they used this an opportunity to discover what approaches were producing the most effective results on a day to day basis.

> Well, with the popsicles sticks what I remember happening is that the teachers who were on that day would end up coming up to whoever was counting popsicle sticks and be like “How did I do? How did they respond?” So it was actually getting teachers really invested in the process and the questions, and interested in the process and hoping that we as an institution, but also they as individuals, had an effect. It was cool. We didn’t intend it to be about individuals so much, but everyone was so in to it and really seemed to like getting that information right away so they could play with it more the next day.

Evaluation encourages consideration of practices

Many people discussed that the very process of identifying TBI questions and how those questions related to underlying program objectives was useful to the department and to

them as individual educators. The construction of simple, specific, and universal targets was a motivation to be more intentional in program delivery.

> It was pretty useful in that it really changed the way we thought while we were teaching. That was one of the big things we noticed. We were thinking about those questions while were teaching and we could target them a bit. And we, as a staff, picked those questions, so they were already things we agreed that we cared about.

Evaluation can and should be ongoing and integrated

Staff spoke of the iterative nature of evaluation and multiple places where questions and answers had led to more questions, which were also explored as the process now seemed part of practice. A number of staff also commented on hopes and plans for upcoming or future evaluation projects.

> We're an action-oriented team, now, trying to do evaluation, and we're still of course asking questions and reflecting on what we're getting and seeing things and we're constantly trying to keep that moving forward and get more knowledge and information.

Additional Themes Uncovered

There were a number of additional compelling and pertinent topics that arose repeatedly in staff descriptions and commentary that were not ascribed specifically to the pre- or post-TBI time periods. These themes are still productively linked to staff attitudes.

Learning about evaluation

While it was not a topic specifically addressed or solicited by either the interview protocol or the survey, several staff commented voluntarily on appreciating the opportunity to learn more about evaluation, the desire to continue that learning, and the idea that they would like to see that learning extend to other members of the organization who have not yet had that opportunity.

> That's how this helped. Learning to be sure we were getting information that was actionable. To ask the questions that would get that. The concept is pretty natural to us, but that was the part of practice that we grew in the most. And I hope that will spread.

The power of the pat on the back

There was plenty of discussion of designing evaluations that transcended the simple confirmation that a program was good or successful, but several staff also commented on how that type of information also felt important and productive as it increased confidence and job satisfaction for educators to know that they were delivering successful and high-quality programming.

> And like, it helps make the activities feel more effective or it helps me feel more like what I am doing is effective. It gives me more confidence that what I am doing is actually effective for my goals.

How TBI Contributed to Change

The themes already explored create a strong indication that the implementation of Team-Based Inquiry as a practice succeeded in inspiring a shift in both attitudes and behaviors and that this implementation was a positive experience for Bakken staff, as stated by one respondent:

> I can't think of another way of doing evaluation that's as worthy, as credible, as useable…as good a fit.

A more specific examination uncovered some of the thematic reasons why this particular method, with its self-similarity to The Bakken's existing practices, was helpful in achieving these shifts.

TBI provided a method to systematize existing evaluative thought processes

This model did not appear to be an outside initiative enforced on staff, but rather, as respondents expressed, an opportunity to make more concrete their already existing informal ways of considering their programs.

> Well I think we have a method that we can use now. If we decide we don't want to assume we're hitting all the right buttons with the teachers, or the students, or the chaperones we know how to make a survey, whether it's online or pencil and paper or popsicle sticks we know how we can gather that information pretty quickly and easily, so it's not reinventing the wheel.

Not only did TBI provide a method, but it advocated tackling evaluation in smaller, simpler, and achievable tasks with an approach that resonated with the staff.

> And then there was TBI and it was like, oh, it doesn't have to be that hard. What if we don't ask everything? You know, it's like good science. You don't test 17 variables at once. Just ask a question or two and figure out a simple reasonable thing to try to give you information on that question.

TBI utilizes the expertise and advantage of program staff involvement

Echoing many of the sources in my literature review, many respondents discussed the value and effectiveness of having the staff who work most closely with the project not only conduct the evaluation, but be around the table to plan what should be evaluated and to analyze and reflect on the data that come in.

> Sometimes it's hard to pull the meaning out, but yeah, I like thinking about how things work in that way. And when it's thinking about how the things I work on work, that's even better.

This feature of TBI was not only effective in engaging the participants to allow for a more discerning examination of a program conducted by those who know it best, but a way of honoring and valuing that knowledge that staff noted and appreciated.

> And I really think that it is good to have us all thinking about what the answers mean. That worked really well, too. Remembering that we know a lot. (Laughs) We're all smarties! And we can figure out a whole a lot if we take the time to look at it and think about it and talk about it. So we should be part of that process, 'cause otherwise all the brilliance is just being wasted!

TBI increases feelings of ownership and investment in evaluation

Many staff expressed that they felt a greater sense of investment or ownership in the process of the evaluation, and thus more focus and intention around its success. This ownership allowed staff to feel more positive about the process of evaluation, but also

more motivated in their participation, as the results were of interest and value to those who were active in the implementation.

> TBI was just the few of us who actually had to teach the program, who actually worked with it, and we were more invested and involved. And so we were able to use what we already knew from delivering the program to really dig at what we wanted to know and write the questions that would do that…that would work for the kids so the whole thing was more…natural to us. And so yeah, we were more invested in the process and more invested in the data.

TBI felt natural/comfortable/logical for staff

There were extensive references by respondents to the idea that not only did TBI feel more comfortable, but that it seemed to naturally stem from their existing thought processes and practices in a way that made sense to them and increased their positive feelings towards the endeavor.

> It was…simple. It just made sense. I think it was what everyone was already doing in their heads. This process was really just about talking about it out loud and in a more formal way, I think.

TBI is like the science we teach

As the respondents are staff of the education department and thus heavily concerned with developing positive and attitudes and practices in science, a number of staff overtly connected the TBI process to the concepts The Bakken espouses in its programs.

> I just think it has given us the opportunity to collect solid information in a scientific way—I have a scientist's background. I like data, solid information you can stand on. We can use that data to look at things as specifically or broadly as we need to. We can really perfect our programs. TBI really is like science, or engineering, like our version of science and engineering at that—honing a program to be its best version, concretely.

While staff did not directly reference WTDS in their explanations (although a number of did casually use the language in their descriptions!), they were still often associating their evaluation experiences with their own teaching. In some cases these connections went even further to associate TBI with the practices of scientists frequently discussed in program curriculum.

> Information. It's all always there, it's just a matter of finding a way to see it for what it is. It's like…like we are always correcting kids that scientists, that Ben Franklin didn't invent electricity, he made discoveries about it. He figured it out. It was always there, he just found the…key, so to speak.

Themes from Drawing Activity

The drawing activity internal to the interview generated varied concepts of evaluation. Most staff chose to express their idea in an abstract form, although there were some who had more concrete interpretations of the question, drawing *an* evaluation as opposed to evaluation, broadly. Interviewees were each asked to describe their image when it was complete.

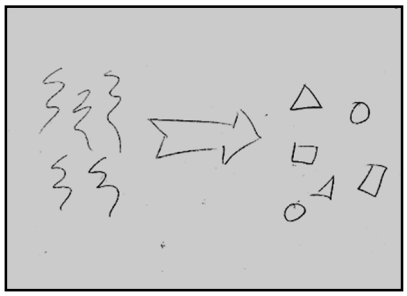


Fig 5.1 Drawing #7

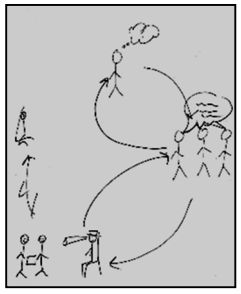


Fig 5.2 Drawing #1

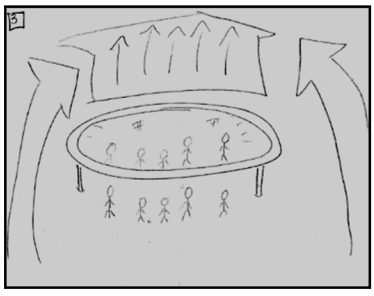


Fig 5.3 Drawing #3

Some of these images took on a likeness to the process or the model itself as in Figure 5.2. This image is reminiscent of both the WTDS and the TBI model graphics. The most prevalent theme in the images was that of arrows. Some of these arrows represented movement through the process while others were described as representing either movement or transformation of information. The interview respondent who drew Figure 5.1, spoke to the trend directly saying:

> I guess arrows make sense here. Evaluation is, well I think it should be…transitive, right? I think that's how we think about it now. Evaluation should be movement. So um, that could be movement of information between people, or movement of an idea, I guess, from one form through its improvements. But it's all for the sake of movement. And that's good! No one wants a stagnant evaluation. (Laughs) We want to ride the arrows!

There was also a somewhat common theme of evaluation serving the purpose of making sense out of complex or confusing jumbles of information. This was best expressed by the respondent who described his image by saying, “From the chaos comes the gold.”

The Future of TBI at The Bakken

Expressed hopes for expansion of TBI from participating staff

Almost every participant, when asked about their ideal for evaluation at The Bakken, discussed hoping to see the spread of these practices beyond the education department. Staff placed value in the practice of TBI and expressed the hope that it can expand throughout the museum.

> And I guess in my hopeful way, if maybe more of this TBI stuff was introduced and if it was done on a more consistent level, then people could be making incremental changes—and see how they are working. Rather than writing this giant survey and saying we’re going to use this to make everything better and actually nothing happens because it’s too big. You can’t look at the pieces for what they are. But if we did like this, this TBI stuff where it was, where it felt like a step instead of a bound it could be something that people could actually accomplish.

When asked to describe their ideal for evaluation, most staff described ideas about not only the expansion of evaluation, but about the move to make it endemic, a process and concept as natural to The Bakken as the Science Assets have become.

> The ideal situation is where...evaluation is just clearly integrated and is there all the time, in everything we do. It's not an afterthought, or a troublesome task to be done. It's something that people find valuable and IS valuable to the organization and the function of the organization. Where people's hard work can be seen, and tracked, and improved through an evaluation system so that it's obvious through people's effort that the outcomes we are achieving are connected and that the improvements that we make are actually improvements.

Still, most staff openly recognized that this type of change organization wide would require significant commitment, momentum, and leadership.

> I think the challenge is I think it would be really important for the Bakken to have a shared sense of purpose about evaluation, or that we move our approach forward so that the approach is institutional instead of departmental.

Indicators of attitudes and mindsets from non-participating staff

The survey, sent to all staff at The Bakken, included several open-ended questions on their perceptions of evaluation and the need for it at The Bakken. Survey participants who did not also identify as TBI participants' responses provided several insightful themes on the possibilities for continued evaluation capacity building at the organization.

Need more evaluation and more people working on it

Through open-ended survey responses, most non-participants indicated an interest in more evaluation taking place at The Bakken, citing reasons such as ensuring the quality of programs, remaining competitive in the market, and recognizing that there is a developing expertise in the staff that could be translated to more programs and departments than solely education.

> As an organization I think we have an overinflated sense of our evaluation activities. From what I see, we do very little and really have not understood how much time and money our savvy competitors are doing. It is an area we cannot continue to partially focus on.

Need for a system to do evaluation

Several responses indicated an interest in evaluation, but a need for concrete structures or systems in place. Associated with the need for these systems, is the need for staff training and motivation to participate in all program areas.

> Programs always need to be re-examined and improved, by having a well established evaluation process we can make sure our programs work as effectively as possible and meet our programmatic goals including meeting the needs of our many audiences.

Respondents also spoke to the need for the commitment of necessary resources such as time, money, and support from leadership to see the process through.

> For evaluation to be effective, it needs to be proactive and part of a master plan, allowing enough time to develop tools, analyze the data, and incorporate feedback before next steps are taken.

Need for new methods and practices

In responses from non-participant staff, there were a number of statements that echoed perspectives shared by TBI participants from the time before they were introduced to TBI. One of these themes was that of repetitive data collection.

> I have been doing evaluation for many seasons now, but I tend to take the same approach, ask the similar questions, consequently I feel the strengths of the insights reaped have diminished.

Another theme reminiscent of the other pre-TBI era themes was that of unused or disappearing data. One respondent spoke quite forcefully of the phenomenon:

> We frequently collect a file cabinet full of data, information, input, and rarely analyze it or use it to make changes. The Bakken is often the place where data goes to die. Information without interpretation is basically useless.

Desire for Teams of Program Staff

Also reminiscent of TBI practice, several staff discussed the desire to have evaluation be done in teams, to have those teams include staff who work on programs in question, and to endow the thoughts and opinions of those staff with value and importance in the process.

> It seems to work best when it's designed and implemented internally by people who know what they're doing. When we hire outside consultants who don't fully understand the program that's being evaluated, we often don't ask the right questions or ask questions about things that we have little to no ability to change results based on answers; when we do evaluation internally, it's often poorly crafted because again, many staff don't have the skills or training to ask good questions but don't recognize that this is a skill they seem to lack.

One comment indicated that being asked to join such a team would be a meaningful message about their opinion in the life of the organization.

> I have never been approached to do so - even the personal encouragement that my voice should be heard would be powerful.

Quantitative Results from Survey

Given that the data from the survey were acquired from a census, there can be no inferential statistics calculated. However, several questions did yield interesting patterns when responses of TBI participants were compared to those of non-participants. One question asked respondents to describe how comfortable they would be conducting each of six skills associated with evaluation on a 4-point Likert Scale from "Not at all comfortable" to "Completely Comfortable." Given the small number of responses, the two negative response categories were aggregated, as were the two positive response categories, in an attempt to make patterns more evident. The chart below shows the percentage of participants of each type who indicated positive comfort level for each activity.

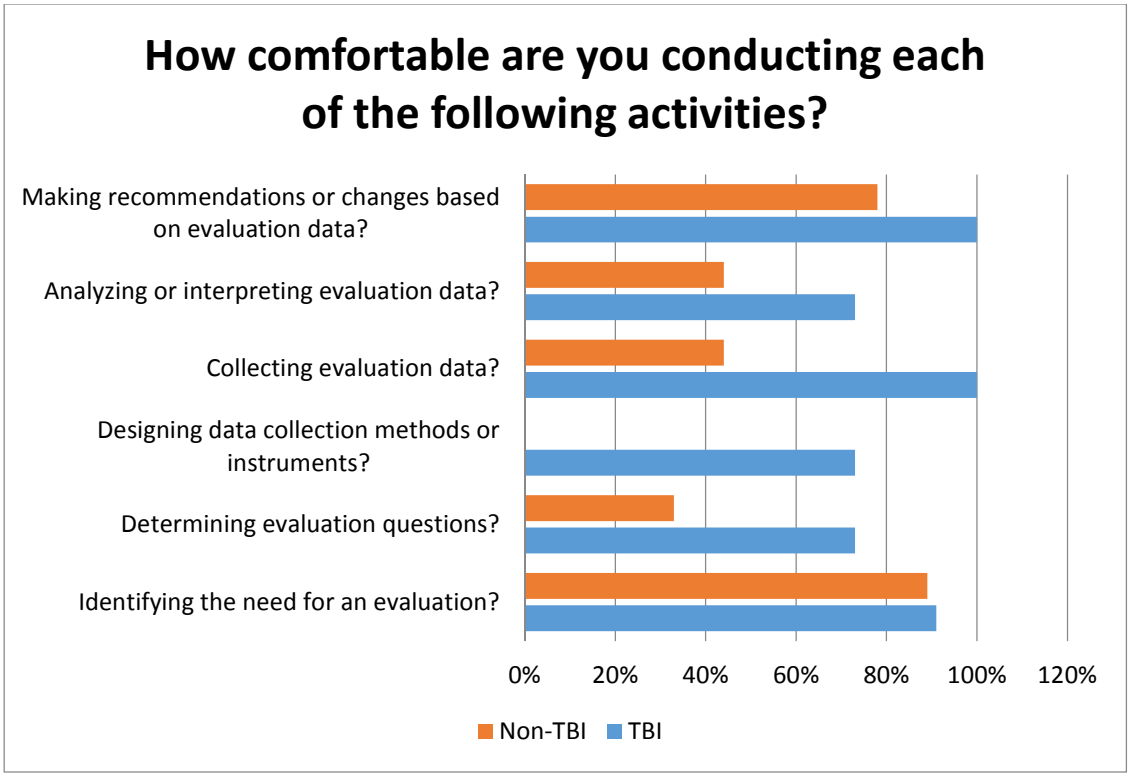


Fig 5.4 Percentage of respondents who indicated positive comfort levels with evaluation activities

The areas that show the most contrast are that more TBI participants appear to have indicated comfort with Determining Evaluation Questions, Designing Data Collection Instruments, Collecting Data, Analyzing or Interpreting Data, and Making or Recommending Changes. A particularly marked contrast is that none of the TBI non-participants reported a positive comfort level with Designing Data Collection Instruments, whereas 75% of TBI participants reported at least some comfort. Additionally, only 45% of TBI non-participants reported comfort with Collecting Data as opposed to 100% of TBI participants. For the category of Identifying Need for Evaluation, there is not a notable contrast. Once again, no inferential statistics can be calculated for these results but, intuitively, they seem to indicate more confidence in evaluation skills on the part of TBI participants than non-participants.

Respondents were also asked to indicate how likely they were to initiate evaluation at The Bakken.

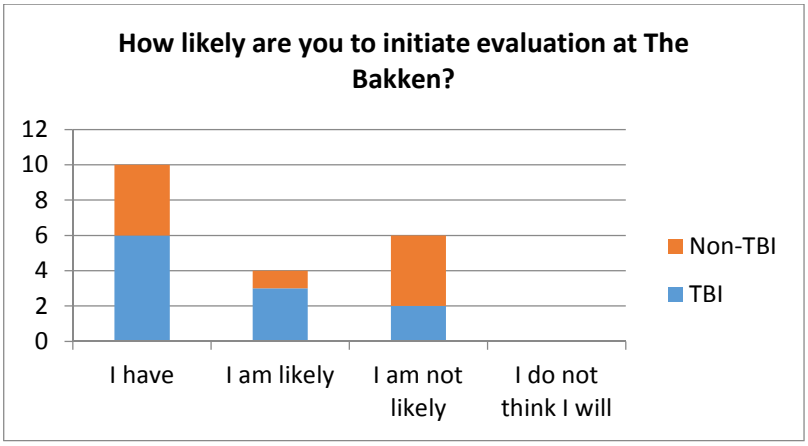


Fig 5.2 Respondents reported likelihood to initiate evaluation

While this answer once again does not provide any definitive evidence, it does demonstrate a pattern of most TBI participants indicating that they either have or are likely to initiate the process, whereas there is a much more mixed response amongst non-participants.

It is also potentially note-worthy that, based on my understanding and observation of TBI activities, of the 30 staff at the museum, only 12 could say they had participated in TBI; of those 12 possible, 11 chose to respond to the survey. Of the remaining 18 staff, only 9 chose to respond. This means that TBI participants' perspective was more evident in the survey, possibly skewing the data, but it is also interesting to note their willingness to respond at a higher rate.

CHAPTER 6--DISCUSSION

What Does It Mean?

A Representative Model of the Data

Taking the primary concrete themes uncovered through the interviews and survey, I have constructed a model (Fig. 6.1) representing the change to staff perceptions about evaluation from before the introduction of Team-Based Inquiry, through its implementation, and to the present. While there are also themes and points of interest regarding the future, I chose to leave them out the model for the sake of clarity.

Representative Model Description

At the center of the model is the idea of evaluation, closely surrounded by the predominant perspectives held by staff towards the practice, as described from the pre-TBI timeframe, in the red ring. These perspectives are then filtered through the concepts espoused by TBI, as articulated by staff, in the blue ring, to then produce the new concepts or perspectives, as they were expressed by staff, of the current status of evaluation in the purple ring. Additionally, there could easily be more arrows continuing outward from there and possibly an interlocking relationship with another series of circles, representing the thoughts and experiences of other departments and, perhaps eventually, other institutions.

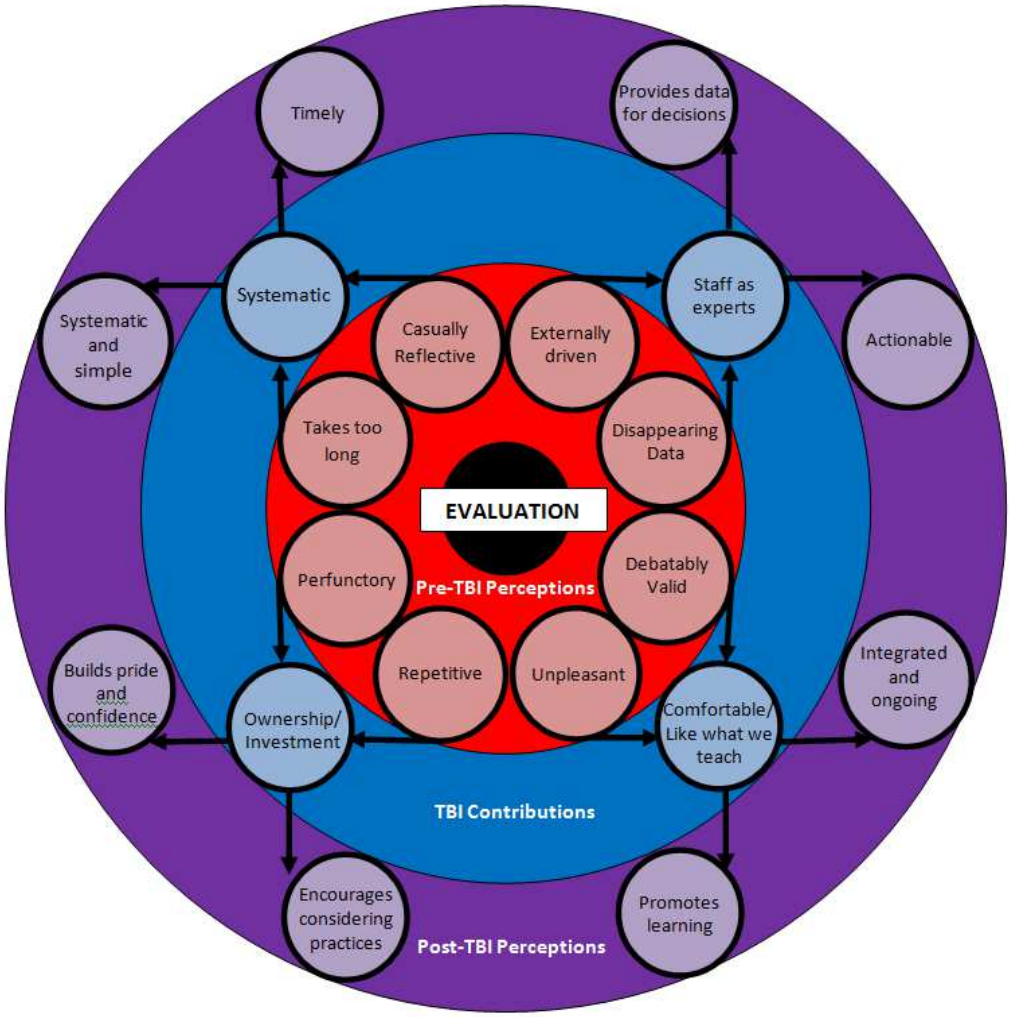


Fig. 6.1 A model of changes to staff perceptions about evaluation at The Bakken

Building Evaluation Assets

The Bakken Museum holds the concept of developing the Science Assets, or the skills and attitudes necessary to succeed in a scientific world, as central to its pedagogy. The implementation of this fractal evaluation model sought to develop the same principles in program staff with regards to evaluation. If the Evaluation Assets can be similarly defined as the skills and attitudes necessary to successfully employ and use program evaluation, there is ample thematic support to suggest that those assets are now present in the staff of the education department at The Bakken. Descriptions of experience with evaluation, use of evaluation, and attitude toward evaluation are predominantly positive (evaluation is useful, logical, attainable, and promotes reflection on practice), but more importantly are practical (evaluation should be actionable, timely, and simple), and enthusiastic about learning and developing more (evaluation should be ongoing and integrated).

Furthermore, there is additional data to indicate that, in many cases, those Assets were significantly weaker before the introduction of Team-Based Inquiry (evaluation was unpleasant, debatably valid, and not used or useful). The qualitative data from this study suggest that this model and the framework with which it was introduced had a positive influence on these important and necessary elements of evaluation capacity building.

How the fractal concept was a factor

In discussing how this new model was able to change their practices and mindsets, respondents revealed multiple themes centered around the comfortable and natural quality of the model, as well as its similarities to the topics taught to students by their own programs. While staff did not frequently report these findings using the explicit

language of Wonder, Try, Discover, Share, suggesting that they perhaps were not overtly aware of the connection, they did speak of the process as being similar and logical to them, and of it reflecting their pre-existing mindset. Further exploration of this idea through another study would be well-served by making the evaluation process more explicitly connected to the taught pedagogy, encouraging the use of that language specifically in framing evaluation practice and making the notion of the fractal or of self-similarity more foundational to the implementation.

Expansion and Dissemination

At The Bakken

As shown through interview and survey responses, staff at The Bakken are hungry for more evaluation. The staff who participated in TBI are largely enthusiastic about continuing the process and also demonstrated pride in the work they had already done, both in evaluation results that demonstrated program successes and in the demonstration of their successes in evaluating the program. The education department, because of these strong assets, stands to continue engaging evaluation and exploring how it can contribute to their work.

There is noticeably more skepticism about evaluation from those who have not participated in TBI than those who have, but there is still interest in the promotion of evaluation to a more regular role in the organization. Many of the negative and wary attitudes these staff members express are similar to those described by staff as their experience before TBI; this would suggest that an expansion of the method to the rest of the organization might be successful.

The education department of The Bakken museum did, however, have a number of features that made it a particularly strong candidate for this implementation. The Science Assets were born of the education department and have been very thoroughly

integrated into all corners of it. While there is movement toward this integration in other program areas of the museum, the direct applications are somewhat less obvious for other fields, such as exhibits or development. In the same way, the fit of TBI might be somewhat less comfortable for those departments. Still, this invitation to curiosity and inspiration is central to the museum's function and mission, and thus the idea is not completely foreign. Furthermore, the other described properties of TBI--simplicity, timeliness, ownership, and practicality should still be appealing to anyone with questions to answer.

Beyond The Bakken

Countless other institutions and organization have very similar educational objectives and could easily adopt an evaluation method that would be similar to their pedagogy. As experts on the field of informal science education suggest, developing science learning, or indeed evaluation learning, should be a process that feels specified to an individual cultural or social context, if longevity of the practice is a desired outcome. (NRC, 2009)

Team-Based Inquiry has enjoyed success in a number of institutions and would, in many cases, be able to be linguistically adapted to fit the described scientific process of any particular program, if necessary. This approach could be particularly helpful in organizations that have not done much recent evaluation or who have historically isolated evaluation apart from the rest of the institution. Negative attitudes about the process seem to be notably enhanced the less transparent evaluation intentions and practices are and the further evaluation resides from program practice and staff, a concept that is supported by contemporary program evaluation standards. (Yarbrough et al., 2011).

The development of these enthusiastic attitudes and inquisitive cultures can also be seen to help when, by necessity, evaluation must be conducted in a more formal or external manner. While the theoretically objective, large scale, external evaluation may no longer hold the premium as the only acceptable form of practice, it is still merited in numerous circumstances. Team-Based Inquiry would not be appropriate for this type of study as it does not necessarily equip participants with the extensive skills needed to manage that level of data collection or analysis. Furthermore, there are instances when the objectivity of an external design is required or called for. Nevertheless, program staff who have strengthened evaluation assets and have integrated evaluation practice into organizational culture might be able to transfer the positive attitudes and provide greater acceptance and assistance to any evaluation, including those that are conducted by external actors. While this type of evaluation approach would not satisfy that need, an external evaluator working with a staff already primed for posing questions and articulating objectives would likely find a these skills, and the associated interest, enthusiasm, and understanding of the Evaluation Assets to be a boon to that process.

Conclusion

As evidenced by the growing prevalence of Evaluation Capacity Building literature, there has been an evolution from the assertion that evaluation must be conducted objectively, formally, extensively and, externally to a more nimble conception, which allows evaluation, when appropriate, to be more customized, enacted and explored by those closest to the object itself. Even with the acceptance of simpler, internal, and less formal practices in the field of evaluation and even with a strong movement to cultivate those capacities, evaluation cannot truly proliferate in an

organization without the perception of it being possible, practical, and personal—the perceptions cultivated by the Evaluation Assets.

The concept behind a fractal evaluation, introducing evaluation as an extension of existing organizational thought, paves the way for the Evaluation Assets to develop by honoring and employing the knowledge and ability of the staff involved. In the education department of The Bakken Museum, TBI was able to provide a model that did that very gracefully; the Evaluation Assets of the staff improved as a result. With those assets in place, evaluation continues to gain strength and momentum as a standard of the practice in that department and will possibly expand into the practices of the organization at large. Additional studies would help to understand how this process might be transposed to a new environment. While this particular case of Team-Based Inquiry and The Bakken Museum is specific and narrow, the endeavor to personify evaluation in this manner appeared to have a positive effect in that environment and could likely have similar effects beyond, in whatever form or format was prescribed in a given situation. If successfully adapted, the notion of fractal evaluation and its ability to develop Evaluation Assets could be helpful to a number of environments seeking to develop a greater capacity to engage evaluation as an endemic practice in their institution.

Beyond increasing the amount of evaluation that will happen in the organization, beyond improving these assets to this practice of evaluation, this evaluation model also uncovered in staff a great deal of satisfaction with their work. While not a specifically documented theme, staff were uniformly pleased with not only the results of their work, but their participation and success in the process of showing those results. The fact that this evaluation process feels as though it is something of their own increases the associated gratification. These staff members are invested, are interested, and are proud of their programs as well as of the things they have done to make those programs

the best they can be. That type of job satisfaction is an excellent side-effect to an already productive practice.

Bibliography

Anderson, K. L. (2013) A retrospective case study of long-term evaluation capacity building at Neighborhood House. Master's Thesis, The University of Minnesota, Minneapolis, MN.

The Bakken Museum. (2015) Home page. Retrieved in March 2015 from www.thebakken.org.

The Bakken Museum (2013) Results from internal evaluation projects. Internal documents.

Boruch, R. F. (1991). The president's mandate: Discovering what works and what works better. In M. W. McLaughlin & D. C. Phillips (Eds.), *Ninetieth yearbook of the National Society for the Study of Education, Part II. Evaluation and education: At quarter century* (pp. 147-167). Chicago, IL: The University of Chicago Press.

Campbell, D. T. (1984). Can we be scientific in applied social science? In R. F. Conner, D. G. Altman, & C. Jackson, *Evaluation Studies Review Annual* (pp. 26-48). Beverly Hills, CA: Sage Publications.

Caracelli, V. J., & Greene, J. C. (1997). Crafting mixed-method evaluation designs. *New Directions for Evaluation, 74,* 19-32.

Cousins, J. B., Goh, S. C., Clark, S., & Lee, L. E. (2004) Integrating evaluative inquiry into the organizational culture: A review and synthesis of the knowledge base. *The Canadian Journal of Program Evaluation, 19(2), 99-141.*

Creswell, J. W. (2013). *Qualitative Inquiry and Research Design*. (EDITION) Thousand Oaks, CA: Sage Publications.

Higa, T.A.F., & Brandon, P.R. (2008). Participatory evaluation as seen in a Vygostkyian framework. *The Canadian Journal of Program Evaluation, 23(3), 103-125.*

Huffman, D., Thomas, K., & Lawrenz, F. (2008). A collaborative immersion approach to evaluation capacity building. *American Journal of Evaluation, 29(3), 358-368.*

King, J. A. (1998). Making sense of participatory evaluation practice. *New Directions for Evaluation, No. 80,* 57-67.

Labin, S. N., Duffy, J. L., Meyers, D.C, Wandersman, A., & Lesesne, C. A. (2012). A research synthesis of evaluation capacity building literature. *American Journal of Evaluation, 33(3), 307-338.*

Mandelbrot, B. B. (1983). *The fractal geometry of nature*. New York, NY: Macmillan.

Marton, F. (1986). Phenomenography - A research approach investigating different understandings of reality. *Journal of Thought*, *21*(2), *28-49.*

Murphy, B. (2012). *Science assets program description*. Internal document, The Bakken Museum.

Nanoscale Informal Science Education Network. (No date). Home Page. Retrieved from www.nisenet.org.

National Research Council. (2009). *Learning Science in Informal Environments: People, Places, and Pursuits. Committee on Learning Science in Informal Environments.* P. Bell, B. Lewenstein, A. W. Shouse, & M. A. Feder, Editors. Board on Science Education, Center for Education. Division of Behavioral and Social Sciences and Education. Washington, DC: The National Academies Press.

Pattison, S., Cohn, S., & Kollmann, L. (2013). *Team-Based Inquiry: A practical guide for using evaluation to improve informal education experiences*. Retrieved January 2014 from www.nisenet.org.

Patton, M. Q. (2008). *Utilization-focused evaluation: The new century text* (4th ed.) Thousand Oaks, CA: Sage Publications.

Patton, M. Q. (2011). *Developmental Evaluation*. New York, NY: The Guilford Press.

Preskill, H., Torres, R. (1999) *Evaluative inquiry for learning in organizations*. Thousand Oaks, CA: Sage Publications.

Rubin, H. J., Rubin, I. S. (2012). *Qualitative Interviewing; The art of hearing data*. (3rd ed.) Thousand Oaks, CA: Sage Publications.

StrateGems. (1997-2012). Natural Hierarchies. Retrieved in April 2013 from http://strategems.com/fractalorgcharts.html.

Volkov, B., & King, J. A. (2007) A checklist for building organizational evaluation capacity. Retrieved January 2015 from http://www.wmich.edu/evalctr/archive_checklist/ecb.pdf

Vygotsky, L. (1997). *Thought and language.* Cambridge, MA: The MIT Press.

Weiler, D., & Stearns, M. (1980). The uses and limits of education evaluation at the state level. In J. A. Pincus, S. E. Berryman, T. K. Glennan, P. T. Hill, M. W. McLaughlin, M. S. Stearns, & D. Weiler, *Educational evaluation in the public policy setting* (Rand Document No. R-2502-RC, pp. 77-85). Santa Monica, CA: Rand.

Wolfram Mathworld (2015) Fractal. Retrieved in March 2015 from http://mathworld.wolfram.com/Fractal.html.

Yarbrough, D., Shulha, L., Hopson, R., & Caruthers, F. (2011) *The program evaluation standards: A guide for evaluators and evaluation users.* (3rd ed.) Thousand Oaks, CA: Sage Publications.

Appendix A

Interview Protocol

The interview should take 30-45 minutes of your time. If at any point you wish to stop or do not wish to answer a question, please let me know and we will move on. Are you willing to participate in the interview?

About two years ago the Education Department at The Bakken first started using our own form of Team-Based Inquiry by applying the Wonder, Try, Discover, Share process to our own programs.

1.Thinking back to before that method was part of the practice here, what are some examples of evaluation that were happening at The Bakken during that earlier time?

>Which of those examples did you work most closely with?
>What role did you play?

2. What was your experience with that/those evaluation/s like?

>What was the goal of the evaluation?
>In what ways did that evaluation achieve its goal? Not achieve its goal?
>What happened with the results?
>How was it useful? Not useful?
>How did you feel about the process?

3. Now, please tell me about an experience with an evaluation at The Bakken that used TBI.

>What role did you play?
>What was the goal of the evaluation?
>In what ways did that evaluation achieve its goal? Not achieve its goal?
>What happened with the results?
>How was it useful? Not useful?
>How did you feel about the process?

4. How did your experiences with the different types of evaluation compare to each other?

>How were they similar?
>How were they different?
> In what ways did TBI, as a method in particular, contribute to those differences?

5. Now we are going to do something a little different. I would like you to take a few minutes to draw a picture of what the idea of evaluation, as a broad concept, is to you. It can be as abstract or concrete as you like, it just needs to fit on one side of one piece of paper. Just let me know when you are done.

Now please explain what you chose to draw
>Where would you put yourself in this picture?
>What does that say about your relationship to evaluation?
>How would the picture change if it were a picture about The Bakken's relationship to evaluation?
>How do you think that relationship has changed since the beginning of TBI?

6. Looking into the future, what would be your ideal for evaluation at The Bakken?

>What are some reasons why that hasn't happened already?
>What are the barriers to that changes happening?

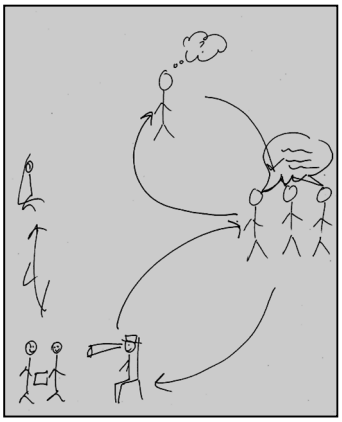
7. What do you think would have to happen for those changes to take place?

>What are some reasons why that hasn't happened already?
>What are the barriers to that changes happening?

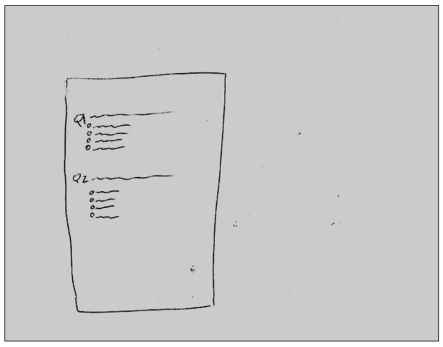
Appendix B

Additional Results of Drawing Activity

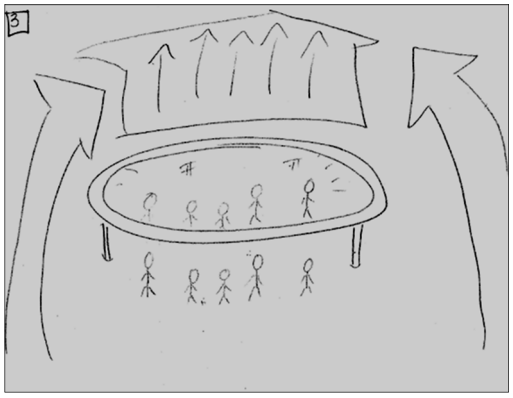
The images and descriptions below were those elicited by the drawing activity included in the interview process. Respondents were asked to draw their concept of evaluation, then to describe/explain their drawing.



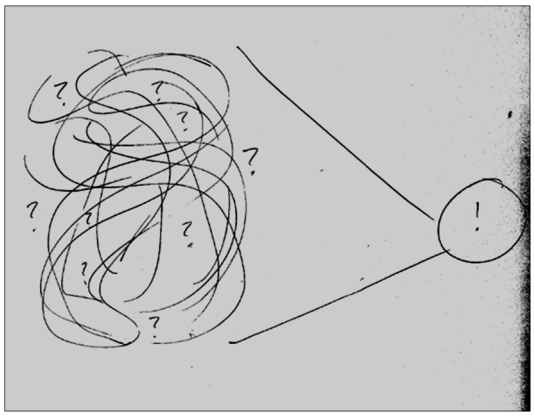
1: Okay so this is a person, if you will, asking some sort question, something that they wanna know more about and maybe they bring to a group of people this group, for some discussion, some brainstorming, things like that. And then they kind of get to take a moment to kind of figure out how are we going to answer this questions. And then come with some form of how they are going to figure out the answer to that questions. Maybe I imagine that form involves something that they are looking or listening or doing something with their senses as they explore possible ideas for answers. I don't know why that one's wearing a hat but he is. And then they talk about it all some more and that leads to more questions. I was gonna draw more people talking some more but I decided just to loop it back and simplify my art.



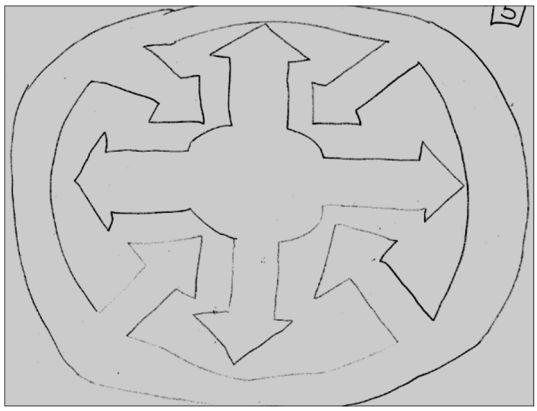
2: So like I said, the first thing that comes to mind is doing some sort of survey. Cause that is often the easiest way to get a little bit of information. And I don't know how to draw observation, other than just binoculars, but that makes me feel like a creeper. But I would put that in there. Um so yeah, so I mean I guess that what comes to mind when I think of evaluation is collecting data. And often we use surveys to do that. So people can share, or we can check boxes as they walk by or whatnot**.**



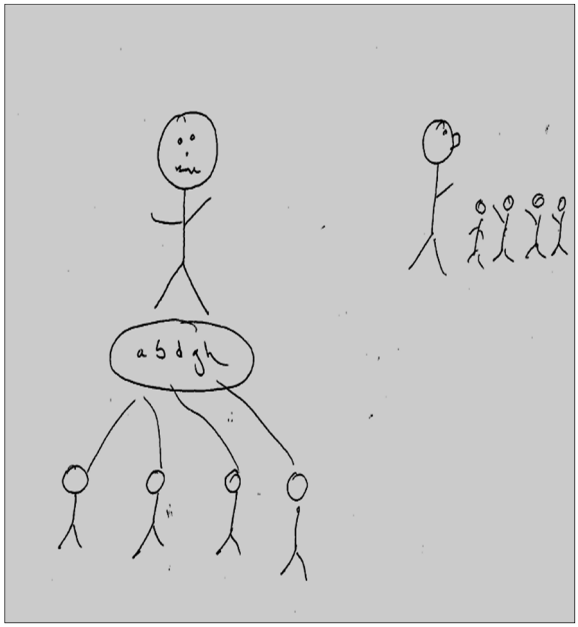
3: So I have some stick figures. And in the center—though it looks like a trampoline—it is intended to be a mirror. So I view evaluation as reflecting the opinions of a group of people. And above the mirror is a large arrow with forward looking arrows. And the intent is that the will, the ideas of the people who have been given a voice been given stake by an evaluation can drive agendas forward. I think the idea is that these side arrows… I think about times that I have used evaluation, like outside of the museum where nothing is all that contentious, and it can be a tool to drive off radical flanks.



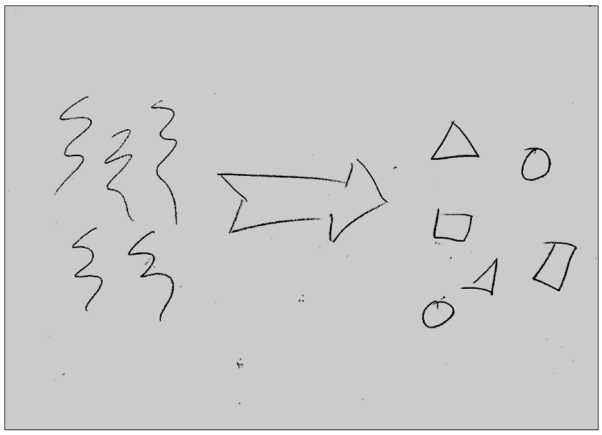
4: I went with abstract. So this is kind of the idea of all the questions that one might have and the chaos of all the different students or whatever it might be and whatever it is that you are trying to assess and then it kind of, with good evaluation it kind funnels down to um good ideas. Yeah. From the chaos comes the gold.



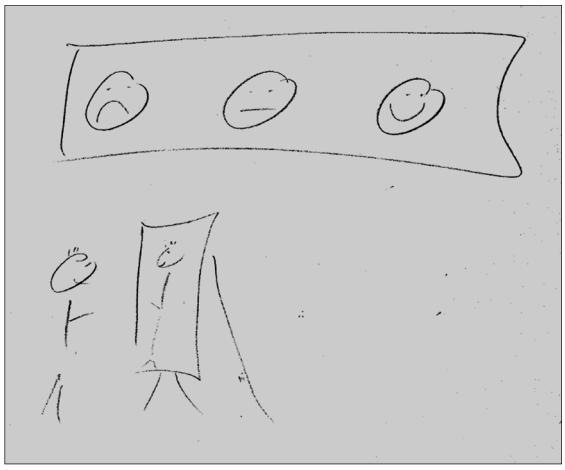
5: I see it as the two-way sharing of information. So I see it as the information here. So this would be us on the inside, that we want to accomplish or to put out into the community. And this is all the information or the experiences that come from outside, the community that we need to gather in here. So that's the information that's coming in through evaluation. So that we can re-evaluate or pat ourselves on the back or say we need to do better, so it can cycle back out. So if this were three-dimensional it would be in motion going out and coming back in and functioning on multiple spheres.



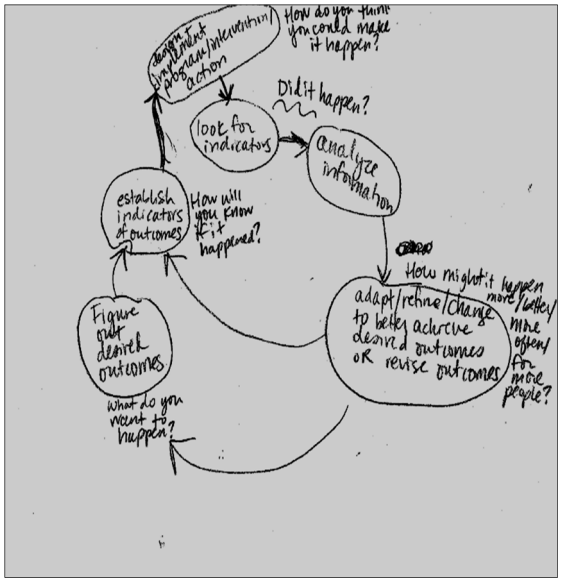
6: It's a little bit two-segmented. 'Cause the first thing I think of, and this might be my teacher training. 'Cause I think of government testing and we just have these tests are being pushed down on kids and we are funneling everything down to these meaningless letters and numbers and making judgments about these individuals. And that's how I've thought about it for a long time. But then I think there's a positive way too. And so I went to the more informal evaluation that as teachers we do all the time. And so I have three kids raising their hands to show they understand you know one is not, so then you know okay I got through to three but one still needs some help. And that's a nice informal way to figure things out. And there's all sorts of range in between I guess. But I am obviously more of a fan of the second way. Yeah, that's all.



7: We are looking at confusion being turned into concrete ideas. You have sort of this unknown just mass, all these squiggles. So you've got an idea what you are doing you have these things that you are using but you're not sure what their effect is or what they're doing and you know that information is there it's just hidden in the squiggle. And with evaluation you can form them into more concrete, useful ideas—that's these shapes. And they are not always completely perfect circles, or whatever, but now that information is in shapes you can work with or understand them better.



8: So what I've done here is describe sort of the continuum that I see in evaluation. One is, as I've seen in the past, happy go lucky surveys. Did it work well for you or not? Those types of really quick, one-syllable answer, type of evaluation metrics. On the other end I have drawn an individual looking in a mirror at him of herself. Which is really a much deeper, metacognitive-based process, by which, with deep introspection we are looking at ourselves and the things that we've been producing to find a sort of deeper meaning in what we are doing, why we're doing it, and the outcomes that we are actually achieving with all the things that we do.



9: So I think the starting point is figuring out what it is that you want to have happen, not necessarily knowing how you would measure it, but thinking about what's the impact you are looking for. Then that going into how would you know it happened So after that is how do you think you could make it happen and then you collect and analyze information and say did it happen. And then going into well how might it happen more better, more often, for more people and that echoes back to this point again where you will have to say how will you know it happened? And so it's a cycle. And it's my belief that you are always going to be able to do it better. Sometimes the change that you make may be more incremental than monumental, but you always can. But it doesn't end. It never really stops.

Appendix C

Survey Instrument Content

This survey was administered online, via Qualtrics, thus the formatting was slightly different than seen below. All content, however, was the same.

Thank you for considering participating in my research. I am trying to learn how The Bakken uses and perceives the practice of program evaluation and how the culture around it may or may not have changed over the past few years. Your thoughts and perspective are very important to helping me understand how this organization as a whole approaches this topic. Participating in this survey or interview is completely voluntary, so if at any point you decide you do not want to, please just let me know or quit the survey, and the process will stop. The survey should take no more than 10 minutes to complete. By clicking next, you will be indicating that you consent to participate in this survey.

Please answer each question to the best of your ability. While most questions are multiple choice answers there are a few short answer questions.

1. Please indicate approximately how many times you have participated in the following activities.

	Prior to working at The Bakken (please select one for each row)			**Since Working at The Bakken** **(please select one for each row)**		
Identifying the need for an evaluation	0	1-4	5+	0	1-4	5+
Determining evaluation questions	0	1-4	5+	0	1-4	5+
Designing data collection methods or instruments	0	1-4	5+	0	1-4	5+
Collecting evaluation data	0	1-4	5+	0	1-4	5+
Analyzing or interpreting evaluation data	0	1-4	5+	0	1-4	5+
Making or implementing recommendations for program improvements or changes based on evaluation data	0	1-4	5+	0	1-4	5+

2. Please indicate how comfortable you, personally, would feel conducting the following activities. (please select one response for each row)

	Not at all comfortable	**A little comfortable**	**Fairly comfortable**	**Completely Comfortable**
Identifying the need for an evaluation				
Determining evaluation questions				
Designing data collection methods or instruments				
Collecting evaluation data				
Analyzing or interpreting evaluation data				
Making or implementing recommendations for program improvements or changes based on evaluation data				

3. Please indicate how comfortable you think The Bakken, as an organization, is with conducting the following activities. (please select one response for each row)

	Not at all comfortable	**A little comfortable**	**Fairly comfortable**	**Completely Comfortable**
Identifying the need for an evaluation				
Determining evaluation questions				
Designing data collection methods or instruments				
Collecting evaluation data				
Analyzing or interpreting evaluation data				
Making or implementing recommendations for program improvements or changes based on evaluation data				

4. Please consider a project or program that you work with at The Bakken that you believe would benefit from evaluation.
    a. Why, in your opinion, should this program or project be evaluated? (please respond in the space below)

    b. How important do you feel it is for The Bakken that this program or project be evaluated? (please select one)
        - o Very important
        - o Fairly important
        - o A little important
        - o Not very important

    c. What types of data collection methods do you think would useful in evaluating this program? (please select any that apply)
        - o Surveys
        - o Interviews
        - o Observations
        - o Other interactive methods (card sorts, word walls, activities)

5. Please indicate how comfortable you would be creating and using each of the following types of data collection instruments. (please select one response foe each row)

	Not at all comfortable	**A little comfortable**	**Fairly comfortable**	**Completely Comfortable**
Surveys				
Interviews				
Observations				
Other interactives (card sorts, word walls, activities)				

6. How likely are you to suggest or initiate evaluation for a project or program at The Bakken?
   (please select one)
    - o I have initiated evaluation
    - o I am likely to initiate evaluation
    - o I am not very likely to initiate evaluation
    - o I do not think I will ever initiate evaluation

   Why do you place yourself there? (please respond in the space below)

7. Considering the above answer, please comment on what could help you be more likely to initiate evaluations in the future? (please respond in the space below)

8. Have you participated in any Team-Based Inquiry (TBI) studies at The Bakken? (please select one)
    - o Yes
    - o No

9. Please share any other comments you may have on evaluation at The Bakken. (please respond in the space below)

    Thank you for taking the time to complete this survey. Your experiences and opinions are very much appreciated!