Evaluating Alternative Dispute Resolution Projects:
Review and Recommendations for the William D. Ruckelshaus Center

Author: Alan Foster
Prepared for the William D. Ruckelshaus Center
June 2011

THANK YOU

Professor Leigh Anderson for being a wonderful advisor
Michael Kern, Director of the Ruckelshaus Center, for your help and guidance
Denis and Bob, loving uncles, for always supporting me

TABLE OF CONTENTS

Alan Foster

Executive Summary

The William D. Ruckelshaus Center seeks to create a project evaluation instrument with which The Center can identify lessons learned, best practices, and foster process improvements.

This paper reviews *program evaluation principles and standards*, *relevant literature* pertaining to evaluating the types of projects The Center engages in, *provides a summary of semi-structured research interviews* of the same, presents analysis of *tradeoffs and benefits* of various evaluation options, *offers recommendations* to fill The Center's evaluation need, and *provides draft evaluation instrument*s which The Center can use to evaluate its projects.

Literature and Interviews:
The literature and interviews validated the need for process evaluation. Facets of evaluation suggested include:

- Evaluating the agreements of outcomes of ADR and collaborative efforts,
- The procedural efficacy and participant satisfaction of those efforts,
- How the relationships of participants are altered,
- The nuts and bolts of process requirements in terms of time and cost,
- What other instruments are currently being used for evaluation,
- What formats of evaluation should be used.

Evaluative Options:
The main options identified for evaluating The Center's projects include:

- Surveying project participants both before and after a project,
- Using in person interviews as well as questionnaires,
- Using a neutral third party observer during project processes,
- Creating post project participant driven focus groups,
- Writing case studies,
- Creating a practitioner forum for professional reflection and sharing.

Criteria for Recommendations:
In order to create recommendations for The Center, options were appraised according to several criteria. These were:

- Compliance with program evaluation principles and theory (reliability, validity, survey construction, etc),
- The Transactional cost to The Center (the pain/gain ratio),
- The degree to which an option meets The Center's needs (lessons learned, best practices, process improvements).

Recommendations to The Center:

- Use a Pre-Post- survey evaluation design
- Use participant interviews
- Focus on four key evaluative elements

Substance – Procedures – Transformation - Outcomes

- Include both participants and practitioners in evaluation feedback

The Center should also consider:

- Gathering feedback from community members not involved in process decision making who are impacted by a project
- Reconvening past project participants into a feedback focus group
- Writing case studies

INTRODUCTION

Alternative Dispute Resolution (ADR) practice consists of processes and techniques which promote collaborative problem solving to assist disagreeing parties in coming to agreement, ideally in a consensus seeking fashion. These processes and techniques widely include mediation, facilitation and negotiation between parties; frequently require convening practices; and practices that include public or community involvement or input. In recent years, ADR processes have taken such a prominent role in public policy resolution that federal agencies have devoted personnel and resources to engage specifically in ADR work.

With the growing use of ADR and collaborative problem solving, there is a growing need to study how basic ADR concepts are evolving and whether on-the-ground implementations are working. While practitioners and academics seek comprehensive approaches to ADR evaluation, field evaluation is progressing in fits and starts. There are few guidelines or agreed-upon criteria on how to evaluate ADR effort, nor on *why* ADR evaluation should be done. However, there seem to be two loose categories of evaluative need, whether or not their import is widely agreed upon.

First, there is need of evaluative evidence that ADR is *working*, though what *working* means is itself a controversy. People who fund or provide political support to ADR projects need proof of the value created ADR projects if they are to continue their support of ADR projects. Parties of interest to public controversies need proof of past success before they become willing to invest their time and patience in lengthy ADR processes. Each of these parties would likely desire slightly different information thus necessitating slightly different evaluative techniques and principles and inquiries.

Second, there is a need to evaluate ADR and collaborative consensus seeking processes in order to find out what the field (and individual projects) is getting right and what it is getting wrong. Because evaluating ADR processes as a whole is ambiguous, and because of the different levels of evaluative need, right and wrong are relative terms. However, they convey the real question regardless of individual feelings on what is right, what is wrong, or if we can be sure. How can we, as a field, improve? Practitioners within the field of ADR do not choose their professions idly.

The William D. Ruckelshaus Center (The Center) is a joint effort of Washington's two research universities and was developed in response to requests from community leaders. Building on the unique strengths of the two institutions, The Center is dedicated to assisting public, private, tribal, non-profit and other community leaders in their efforts to build consensus and resolve conflicts around difficult public policy issues. The mission of the William D. Ruckelshaus Center is to act as a neutral resource for collaborative problem solving in the region. The Center provides expertise to improve the quality and availability of voluntary collaborative approaches for policy development and multi-party dispute resolution.

This paper aims to create an evaluation instrument for the William D. Ruckelshaus Center that can used on a project-by-project basis to capture lessons learned, foster best practices and continually create process improvements. This instrument is not created with the intention of satisfying funders or political authorizers, however The Center can employ instrument results as The Center sees fit.

PROGRAM EVALUATION

Program evaluation traditionally asks two types of questions: Summative evaluation asks questions such as "does the program work." Formative evaluation asks questions such as "Is the program constructed appropriately to achieve its desired goals." From the start, case by case project evaluation faces several challenges. It is near impossible to conduct a randomized experiment, because no two disputes are alike and no two ADR projects are alike. Case-by-case project evaluation of ADR processes cannot achieve quasi-experimental status, either, because there is no control group to be had. Because of these limitations it is difficult for ADR evaluations to offer substantive inference and attribution. Yet another is that project evaluation often seeks to answer both formative and summative questions simultaneously, which causes significant design and usability hurdles.

However, this is not an impasse. Large-scale evaluations of ADR as a form of intervention have relied on broadly gathered empirical research. There have been successful theoretical comparisons made with control groups and control processes (alternatives to the ADR process such as litigation). This is beyond our current purpose. The purpose of creating a project evaluation instrument for the Ruckelshaus Center is to provide The Center with case by case feedback so that they may capture lessons learned to continually improve their project processes.

Nevertheless, in order to make sure that the project evaluation instrument is sound, I consider several standard program evaluation principles and strive to avoid certain pitfalls. I first briefly describe the principals and pitfalls. In the Tradeoffs section I analyze instrument style, instrument implementation, and instrument question-types accordingly. In addition, I will analyze the aforementioned three areas of instrumentation in terms of transactional cost, the amplitude with which they can achieve the Ruckelshaus Center's project evaluation goals and in some cases the usability of the instrument.

Program Evaluation Principles

The project evaluation instrument that this inquiry will produce must take many issues into consideration in its design in order to maximize its usefulness. Typically in program evaluation there are more issues to consider than the following: these are the most primary to a project evaluation instrument for the Ruckelshaus Center.

Reliability

In the case of a project evaluation, the reliability of an instrument is the extent that an instrument produces the same results when assessing one measure multiple times. For example, a thermometer that returned different readings with multiple contiguous uses would be an unreliable indicator of a fever. In that light, an evaluation instrument must contain a degree of specificity that minimizes rapid changes in interpretation in the individual responding to the evaluation inquiry.

Validity

Validity in evaluation is the degree that a measure captures the information it is intending to capture. In the ambiguous category of an alternative dispute resolution process's success, what success means, or what various outcomes could be reasonably thought of as successful, would have to be clearly spelled out ahead of time. If not, trying to capture the degree to which a process succeeded would be left to the mercy of interpretations that could be contrary to the processes intended purposes.

Sensitivity

Sensitivity is essentially the level of nuance that a measure is able to incorporate into its findings. For example, a project evaluation seeking to understand the level of participant's satisfaction asking only if participants were 'satisfied or not' would results in potentially superficial results. On the other hand, if that same measure asked participants to report their level of satisfaction with specific activities, it would generate more detailed and usable information. Another example is if an ADR process was interested in transforming relationships between participants, then the evaluation of the process would have to be able to capture the nature of participant's relationships before and after the ADR process, and it must be sensitive enough to discern these changes.

Biases to project evaluation

Unintended Consequences

In general, project evaluation must be mindful of creating effects that the evaluation does not seek to create. For example, an evaluation of a dispute resolution process implemented in an untimely manner could create political pressures among the stakeholders. An evaluation of the skill of the facilitator could fall into the wrong hands and create negative consequences for that facilitator's career.

The Hawthorne Effect

A famous unintended consequence of evaluation is the Hawthorne Effect, in which participants of the evaluation alter their behavior as a result of their knowing they are being evaluated.

History Biases

History biases occur when an event outside of the project impacts project participants. For example, if a stakeholder group in an ongoing land use dispute secures a federal grant during an ongoing mediation, they may suddenly be amenable to an entirely different range of financial outcomes. Their increase in ability to come to agreement on financial fronts was not caused by the mediation process, so an evaluation of the mediation should not attribute all of this group's change in attitude to the mediation process. Because of this a project evaluation has to separate outside from inside events, and weed out irrelevant information. It needs to be constructed in such a way that captures only process relevant information. Otherwise the evaluation will create reports based not on the process it seeks to evaluate, but on causes far beyond the process scope.

Maturation Biases

People continue to grow and learn and age through a project. A Maturation issue is when a participant gains skills outside of the project being evaluated. A basic example: if a program were to attempt to increase the math skills of a young person, the pre and post evaluation may capture not only the benefits from the math program, but also the normal course of childhood intellectual development. A participant in a facilitated consensus-seeking dispute may become curious and decide to read books on consensus building. That participant will gain skills that aid them in the facilitation from a source outside of the facilitation process. Because of this an evaluation instrument that sought to capture a participant's gain in negotiation skills as a result of the facilitation process would potentially misattribute its findings.

Instrumentation Biases

If a neutral third party observer is used to evaluate a process, that observer may become tired or bored over time. In this case the instrument alters during the process and the results produced are altered as well. In the case of a pre and post evaluation, if changes are made to the evaluation instrument in the interim, the results produced could be attributed to the altered instrument rather than the project being evaluated.

Dropout Biases

Dropout threats are very important to project evaluation for The Center. If some number of participants in a Center project quit halfway through, and only those participants who complete the process participate in the evaluation, then a very important segment of stakeholders are not represented and the evaluation results become skewed. This is important as The Center's purpose in evaluation is to create lessons learned and foster process improvements: project participants who drop out are a highly valuable source of information and feedback on what went wrong.

Reasonability Biases

If the project evaluation relies on responses from participants, then what the participants are asked must be within the realm of their ability to respond. Overly complex or theoretical inquires, though interesting to the analyst, could be unanswerable by project participants.

Selection Biases

Since project participants submit data voluntarily, participants self-select both who will respond and what data researchers will and will not receive.

Transactional Cost

Transactional cost is a measure of the time, effort and resources required by any method of project evaluation weighted against the quality of information produced. If the project evaluation instrument requires a great deal of time and effort from either The Center or respondents, the quality and/or completeness of information gathered may be compromised. For example, electronic resources like Survey Monkey are inexpensive, but may result in low response rates. A resulting data set may be too small to be useful. If the instrument is administered by hand via one-on-one interview techniques it may generate a high quantity and quality of information, but at high cost in terms of time, budget and effort for both The Center and the interviewee.

Usefulness of Information

Because of reasonability and transactional cost, The Center faces some constraints. Project evaluation cannot be so large or complex that evaluation participants are unable to provide credible feedback. Further, The Center has specific purposes for evaluating its projects: as stated, lessons learned and process improvements. The Center's project evaluation instrument must be mindful of this and use the *types* of questions and *method* of evaluation that help respondents to provide good feedback that directly helps The Center meet its goals.

METHODOLOGY

In order to understand:

- Current ADR evaluation theory and practice,
- Evaluation options and processes,
- Guiding evaluation principles.

The research arm of this project entailed:

- A literature review,
- Soliciting expert opinions and views,
- Reviewing existing evaluation instruments.

For expert opinion, 25 semi-structured interviews were conducted with ADR practitioners, academics, and participants of past Center projects. Interviews were conducted in person and over the phone, and in one case via email. Interviews were confidential and no attributions are made during the Interview Review chapter. Interviewees were made specifically aware of the scope and intent of this project and the instrument that it culminates in. The purpose of the interviews was to gain insight into what professionals, academics, and project participants view as the necessary components to evaluating the type of projects The Center engages in.

Several currently used ADR project (or process) evaluation instruments were reviewed. These include instruments come from: the Environmental Protection Agency (EPA), the U.S. Institute for Environmental Conflict Resolution (USIECR), Oregon Consensus, and the Rand Corporation; Timothy Hedeen, PhD.; Franklin Dukes, PhD.; and several used by practitioners that were created in-house and are confidential. Note: The EPA and Oregon Consensus use adopted versions of those created by the USIECR. Some of these instruments, and some ADR evaluation theory, rely heavily on deductive frameworks or logic models in order to create evaluations. Reproducing that approach is not the goal of this paper. The purpose of looking at currently used evaluation tools is to identify what practices are currently used and if there are any universal themes among them.

The analysis portion of this project relies on program evaluation standards, principles and theory using mainly Michael Quinn Patton's Utilization Focused Evaluation, and Peter Rossi's Evaluation, A Systematic Approach.

Literature Review

The literature review aims to understand the development of ADR evaluation within the field as a whole. The purpose of the instrument developed for the Ruckelshaus Center is specific to creating process improvements in their public policy ADR projects; however the literature review portion of this paper takes a slightly broader view in order to comprehensively capture the field's general sentiment regarding evaluating ADR processes. The literature review relies on peer reviewed journal articles which were found via the online journal databases at the University of Washington Library (jstor, etc), the Wiley website, and independently published relevant literature. Journals include *Conflict Resolution Quarterly* and *Public Administration Review*. Other sources of relevant literature include the RAND Corporation, the U.S. Institute for Environmental Conflict Resolution, the Federal Register, and *Consensus*. Other journals with relevant articles and selected books were also useful. It should be noted that the amount of literature published on ADR evaluation is not vast, which is itself informative.

Guiding Questions

The literature surrounding evaluation of ADR projects is small. It is even smaller when the evaluation is targeted to a specific area or for specific goals. The following two sections touch on multiple reasons why one would engage in evaluation, however the purpose and question of the current inquiry is: How can the William D. Ruckelshaus Center evaluate its projects to identify lessons learned, understand best practices, and foster continual process improvements? The aim of this project is not to legitimize ADR to its detractors or the public at large, satisfy policy makers, help academia better understand the world, or prove to the funders that their money is well spent. The aim of this project is to research and create an evaluation tool so that The Center can better achieve its mission by continually reflecting on, and improving, its collaborative projects.

What does the literature say about:

1. Why should evaluation be done?
2. Definitions of success of collaborative public policy ADR processes?
3. What should ADR process evaluation measure?
4. How should we implement evaluation?
5. Cautionary messages to be aware of?

Why do Evaluation?

What should the evaluations be used for?

A couple papers responding to this question suggest that reports should be used by the neutral who worked on a case to help him/her improve their practice (EPA 2008, Church 2002, Rolph 1995), or that Funders may want to see a broad array of projects evaluated with aggregated[1] results to ensure that money is well spent (Church 2002). The evaluation could also look at the relationship between

[1] Aggregated results showing statistical significance demonstrate that the field as a whole is succeeding thus creating reassurance for funders and reducing the need for project by project monitoring.

expected outcome and case indices and statistics (time of process, number of participants, hours spent at meetings, etc) (Orr 2008), and thus tell us what sort of track we are on. Larger empirical studies have been used to either prove or disprove the value created by ADR projects, frequently to defend or attack the efforts of the work itself. Lastly, evaluations are often used to justify the existence of a program to a body of authorizers. Essentially, evaluation asks 'is it working' and 'what do we need to do to improve it'.

For environmental conflict resolution (ECR), Frank Dukes (2004) suggests that perhaps the most compelling question evaluation asks is how ECR compares to other ways of addressing [environmental] conflicts. This question applies to a spectrum of ADR projects broader than environmental. Analysts of ECR, proponents and critics alike, commonly think of it as an alternative to other, more traditional, processes. Traditional procedures primarily include legislation, administrative decision making (agency rules and regulations), and adjudication. The risk in thinking of ECR [and ADR] as only an *alternative* to 'Traditional procedures' is that we fail to appreciate that ECR processes are less often *alternatives* and more often one part of a complex and interdependent system of legal, legislative, or administrative processes (Dukes, 2004).

Conley and Moot (2003) echo many of these feelings while saying that motivations for evaluations vary. Participants may want evaluations to help them improve their efforts, facilitators may want to improve their practice and identify what sort of project processes are best for different types of conflict scenarios, policy makers may want evaluation feedback to help create regulations, authorizing bodies desire proof of the value of ADR processes, legislative or agency mandates may have performance measurement requirements, while funders may want to be assured that their money is well spent. Early evaluations of environmental mediation, for example, focused on measures addressing cost, time, fairness, innovation, and longevity of agreement. Further considerations for evaluation include critics wanting to demonstrate the legitimacy of their concerns and academics may use evaluation results to help explore all of the above (Conley and Moot 2003).

What does Success in ADR projects mean?

The definition of 'success' in consensus seeking ADR processes is an area if inquiry which is frequently disputed. Some feel that success should be seen in terms of tangible outputs and outcomes, for example time and cost avoided (Sharf 2004), while yet others feel that aiming for pre-defined success undermines the transformative nature of these types of processes (Dukes 2004). Moore suggests that participants view success in four categories: political acceptability, interest satisfaction, participant ownership, and improved relationships (Moore 1996). Many feel that there is no set-in-stone definition of success. (Church 2002, Orr 2008).

A failed process, perhaps one with no agreement reached, can have lasting effects on the participant's ability to work collaboratively. Participants in mediations that do not come to a conclusive agreement nevertheless derive significant benefits from the mediation. Those potential benefits include parties identifying or discovering their own real interests, generating new ideas for solutions, providing insights for regulators, and improving negotiation skills (Dukes 2004).

A process that reaches very solid and long lasting agreements, including monitoring and party accountability, may fail on the grounds that participants end the ADR process more hostile toward each other than when they began, while being bound by ADR process outcomes. Further, how success is seen can depend on the relative perceptions of participants and external audiences, and can be relative to types of conflicts themselves. A twenty year long hostile environment could be a success simply because the parties are willing to come to the table. Attempting to capture and generalize any one measure of success across multiple projects and stakeholder groups runs the risk of ignoring the important facets of individual areas of conflict and resolution.

The U.S. Institute for Environmental Conflict Resolution said in 2008 that their evaluation tool was developed without any particular benchmark for success in mind. Instead, the audience could decide for themselves what success in ADR meant and then use evaluation results as needed.

Different perspectives on success could include:

- Stakeholders: Achievement of the process goals as defined by the group at the outset.
- Agencies: Completion of mission critical project goals. Improved working relationships with stakeholders.
- Practitioner: Stakeholder satisfaction with process, stakeholder endorsement of ADR [ECR].
- Sponsor: Effective use of resources relative to alternatives

(Orr 2008)

In many ECR processes that "fail" in outcome terms, responses to participant satisfaction surveys nevertheless assert strong elements of "transformative success." Participants speak of "gaining a better understanding of the other parties' interests and perspectives" and "breaking down stereotypes of each other as we spend time face to face and listening to each other" (Foley 2004).

However, program effectiveness can tell us how a program design and administration can be improved (Sharf, 2004). In general, measures of success have a stronger chance of holding ground if they are focused on what value is created by the ADR process. We should keep in mind that a public policy ADR project is often one segment in a larger process, so identifying the 'when' in the project that success should be captured ignores the long term cycle of policy conflict and resolution. Lastly, the conflict surrounding the definition of success should not suggest the absence of success, but instead bring to light that there are varieties of success.

What Should We Be Measuring?

Figuring out what to measure is indeed a daunting task. Evaluations take different shapes depending upon an evaluator's needs (Conley, Moot, 2003). However, even when measures come up short there can still be positive and long lasting social consequences among participant's knowledge and understanding and working relationships with each other (Buckle, Thomas-Buckle 1986).

According to Conley and Moot, 'the most common form of evaluation focuses on whether and how collaborative efforts meet their identified goals and objectives'. In goal based evaluation, process outcomes are measured against goals that are stated at the outset of the project. Because of the variety of projects that come under the scrutiny of The Center, program goals can come from multiple sources including legislative or agency directives, or from early stages of group collaboration. Two other types of evaluations are summative evaluations that seek to understand the outcomes or impact a program or project has, and formative evaluations that seek to understand how a program or project is operating and generally attempt to identify program or project improvements. Standard evaluation practice and theory of consensus seeking ADR processes tend to do more than one of the above simultaneously.

Tangible Outcomes and Traditional 'Hard Measures'

Outcome evaluations compare actual outcomes with desired outcomes. These may be social, economic, or behavioral (Conley and Moot 2003) and typically consist of 'traditional measures'. Traditional measures (or standard measures), that I refer to as 'hard measures', are frequently whether or not agreement is reached, the extent to which agreement was reached, the durability of agreement, and participant satisfaction expressed either by those names or their functional equivalent (Conley and Moot 2003, Orr 2008, EPA, Emerson 2009, Sharf 2004).

"As the number and types of collaborative activities have grown since the 1990s, so have the challenges of understanding the design, management, and performance of collaborative arrangements." (Koontz, Thomas 2006) While Koontz and Thomas as speaking more directly of environmental outcomes, this question applies to broader considerations in public policy collaborative ADR projects. Further, Koontz and Thomas are also particularly interested in *outcomes*. Outputs are the plans, projects and other tangible items generated by collaborative efforts, outcomes are the effects of outputs on [environmental] (here I would substitute 'dispute environment') and social conditions. *Social outcomes*, they say, include trust, legitimacy and social capital, and economic conditions.

Participant Perceptions, Soft Measures and Process Measures

A very common measurement category is participant perceptions (Conley and Moot 2003, Sharf 2004, Orr 2008, Hedeen 2002). This type of measurement is captured typically with a post process survey or questionnaire and asks participants to reflect on the process and report on their experience. It may outright ask them for their 'perceptions' or it may simply ask for their feedback; that the feedback consists of their perception is implicit.

Other considerations of desired outcomes, which I'm referring to as 'soft measures', include results that appear fair to the community, improve relationships among the participants, and maximize joint gains (Buckle and Thomas-Buckle 1986). Other soft measures include how well the parties communicate, learn about each others interests and positions, increase in participant's ability to collaborate effectively, etc. Soft measures can be seen as indicators that point to desired hard measures. Emerson believes that "Effectively engaging the parties certainly appears to be a gateway factor and key predictor of agreements reach, their quality, and improved working relationships" (Emerson 2009).

Soft measures regarding process have been identified as information participants received about the process;

- Their ability to present their side of the dispute
- Amount participants participated in the process and were able to communicate (Susskind & Field, 1996)
- How much control participants had over the process; whether or not the 'right' people were part of the process (Bingham, 1986, Todd, 2001)
- Whether or not the examination of technical and scientific issues was made clear and accessible to participants (Susskind & Field, 1996)
- Whether process participants gain knowledge as a result of the process (which can be shared with others) (Innes, 1999)
- Whether or not participants felt fairness in the process (O'Leary 2001).

Participant evaluations are used to identify stakeholder attitudes, opinions, and relationships; the reduction of conflict between parties; increases in social capital; and other social changes (Conley and Moot 2003).

Other measures of interest include whether or not the ADR process was more effective than participant's best alternative, if participants are satisfied and endorse the process, benefit over cost is appropriate, some kind of public benefit is obtained (Orr 2008). Additionally, neutral 3rd parties are often also asked questions about process and outcomes.

To recap, the literature touches on several types of measures:

- Hard measures that focus on tangible outcomes; measures that rely on participant perspectives
- Soft measures that focus on transformative measures
- Measure that focus on the nature of the collaborative process
- Following several examples of frameworks currently used to evaluate ADR projects.

Examples of Evaluation Frameworks

In general, evaluations seek to impose a set of criteria (the measures) retroactively onto a social event in order to extract information about that event to be used for the various purposes listed previously. The following examples are frameworks for (broadly) the type of evaluation The Center wishes to engage. They are not precisely lists of questions posed directly to evaluation participants (and/or neutral 3rd parties). They are categories and subcategories of question types. These models can inform the crafting of actual evaluation questions.

Criteria Focused Framework

Conley and Moot (2003) specify that any evaluation is based on comparing reality to a set of criteria. They present the following as typical evaluation criteria, but warn that criteria used for evaluation needs to closely match the evaluations goals. They also offer a third category which is specific to the outcomes of the individual project area (IE environmental, transportation, etc.).

Process criteria

- Broadly shared vision
- Clear, feasible goals
- Diverse, inclusive participation
- Participation by local government
- Linkages to individuals and groups beyond primary participants
- Open, accessible, and transparent process
- Clear, written plan
- Consensus-based decision making
- Decisions regarded as just
- Consistent with existing laws and policies

Socioeconomic outcome criteria

- Relationships built or strengthened
- Increased trust
- Participants gained knowledge and understanding
- Improved capacity for dispute resolution
- Changes in existing institutions or creation of new institutions

Logic Model Framework

The U.S. Institute for Environmental Conflict Resolution, in a join effort with a large group of ADR professionals, created a logic model to identify key process components that could lead to desired outcomes (Emerson 2009, Orr 2008). Two examples of evaluation frameworks structured as logic models are:

Logic Model #1

Inputs *in terms of participation and capacity*:

- Was this the right process?
- Did parties have capacity to engage in the process?
- Did parties understand issues and areas of disagreement?
- Was the best information available?

Activities:

- Were parties are able to communicate effectively?
- Were issues that parties cannot agree on are addressed with other approaches?

Outputs/Outcomes:

- Did parties reach complete and durable agreements?
- Were parties are satisfied with what they have achieved?

Logic Model #2

Inputs:

- Was there a case assessment?
- Was an ADR process deemed correct?
- Were the correct parties involved?
- Did parties have time, skill and resources to engage?
- Was an appropriate mediator selected?
- Did the mediator's skills and practices add value?
- Was correct high quality information is used and shared?

Activity:

- Were participants actively engaged and are communicating?
- Did parties know how to narrow issues in dispute?
- Did parties understanding of issues improves?

Outputs/Outcomes: Three primary outcome areas: Reaching agreement, quality of agreement, improved relationships among parties (Emerson 2009).

- Are there good prospects for implementation?
- Were participants satisfied with the mediator?
- Was agreement reached?
- Was agreement is high quality;
- Were relationships are improved.

(Emerson 2009, Orr 2008).

Participant-Centric Model (Hedeen 2002)
Timothy Hedeen, Phd., offers the following four categories for measuring the efficacy of ADR processes that he presents as focused on capturing participant perceptions and experiences:

Program Efficiency
- Cost to participants
- Time from referral to resolution

Program Effectiveness
- Outcomes of mediation
- Participant satisfaction with mediated outcomes
- Durability of mediated outcomes
- Impact on relationship between participants
- Program neutrality

Mediation Process
- Appropriateness/Usefulness
- Preparation process and materials
- Fairness *(opportunity to tell story, feeling understood, respectful treatment, control over outcomes)*

Mediator Performance
- Skills of the mediator
- Knowledge of the mediator
- Impartiality of the mediator

Lee Sharf (2004) offers a list of measures, which he ties to measures of success, that combines all of the types of measures outlined above. Sharf has much to say on each item listed below, for brevity I've included only the topical focal points. They are:

Efficiency
- Cost and Time.

Effectiveness
- Dispute Outcomes and Durability of Outcomes.
- Rate of dispute recurrence.
- Negative impacts.
- Management perceptions and Public perceptions.

Customer Satisfaction
- Participants' Satisfaction with Process, and perceptions of Fairness, Appropriateness, Usefulness, and control of decision making.
- Impact on Relationships Between Parties.
- Nature of relationships among the parties: Are they changed?
- Participants' satisfaction with outcomes.
- Participants' willingness to use alternative dispute resolution in the future.
- Would participants elect to use alternative dispute resolution in future disputes?

Program Quality
- Participants' perceptions of the appropriateness of staff and user training.
- Do participants feel that they were provided with sufficient initial information and/or training on how to use the program?
- Do participants feel that program staff had sufficient training and/or knowledge to appropriately conduct the program?

Participants' perceptions of competence (including appropriateness of skill levels/training):
- Do participants feel that neutrals were sufficiently competent or trained?
- Do participants feel that more or less training was needed?
- Participants' perceptions of neutrality/objectivity.
- Do participants feel that neutrals were sufficiently objective?
- Do participants feel that neutrals were fair in their handling of the dispute?

When comparing these five examples there are several items to note. First, there are no universally overlapping questions or *types* of questions, despite the prima facie similarity of the examples. For example, combining questions directed at clarity of process and usefulness of information (useful information can be thought of as substantively clarifying) yields a universal category only by allowing 'preparation process and materials' from Hedeen's Participant-Centric Model. Second, despite the five example's differences, there are many types of questions where are asked in at least two. Third, each example is derived from deductive reasoning.

How are evaluations implemented?

Probably the most common information point for gathering data about ADR processes is from participants themselves. Participants are given surveys, questionnaires, or asked to participate in semi-structured interviews (Conley and Moot 2003, Orr 2008). Questionnaires may be distributed in person, by mail, or online. Each approach has its benefits and drawbacks and surveys are attitudinal in nature and typically use an ordinal rating scale (Hedeen 2002).

The U.S. Institute for Environmental Conflict Resolution uses surveys administered by a program manager and entails several different types of surveys (depending on the project) for both practitioners and participants. It's been speculated that social methods (surveys, interviews) are best used to capture project outputs while social methods combined with empirical research are better suited for measuring outcomes (Koontz and Thomas 2006). Case studies seem to have fallen out of favor a number of years ago, yet they are favored by many researchers as they permit extensive analysis (Conley and Moot 2003).

As there is debate about all aspects and nuances of ADR evaluation, there is debate about who should be deploying the evaluation. Judith Innes has indicated a call for third-party evaluations to avoid bias in results and avoid the problem that many ADR practitioners face: they don't know much about evaluation principles. However, others feel that the practitioner sits in an advantaged position to perform evaluations as the practitioner has detailed knowledge of the subject material, project history, and (hopefully) has built trust with the parties. Yet another form of evaluation deployment entails participants being in charge of evaluation themselves wherein they conduct self-evaluations and focus groups as part of the ADR process itself. Another, yet time and cost intensive method, uses a neutral observer as part of the project process (Rolph 1995).

Other methods of evaluation include focus groups, practitioner round tables, professional evaluators for meta-review, and employing multi-phased project evaluation at key milestones in a project's development. A before-and-after design offers a practical and deployable solution in these projects.. Further, to ensure that we do not capture bias from one single *phase* of a project, evaluation should be iterated at different points of a consensus building project (Rolph 1995, Church 2002).

Hedeen suggests the following for implementation guidelines for data collection:

- Allow for collection of information on a wide range of topics,
- Require a relatively short time for respondents to complete,
- May be completed at times convenient to the respondent,
- Do not require extensive training of staff or mediators to administer,
- Allow for translation into alternative formats and languages,
- Provide data that can be easily collected and analyzed, and
- Provide findings that may be summarized and presented clearly.

The literature provides information about the current thinking surrounding ADR process evaluation. This perspective is complimented by the insights captured in the following expert and participant interview summary.

Summary of Interviews

Interviews were conducted with 25 ADR practitioners, academics and Ruckelshaus Center project participants. Interviews were confidential and used a semi-structured snow-ball interview structure. Each interviewee was informed of the project's background, scope and purpose. Interviewees understood that they were being asked questions to create a tool that might improve the ADR process, rather than information for the legislature or program sponsors. Interview questions were written in collaboration with Ruckelshaus Center leadership. Interviewees were asked to share their thoughts, ideas, and perspectives on the following issues.

1. What is the most important thing to learn from an evaluation of a project aimed at fostering collaborative public policy?
2. What are the best indicators of a successful or unsuccessful project that project managers could look for and participants could report on?
3. What specific questions or areas of exploration did interviewees feel *must* be included in a project evaluation instrument?
4. What current evaluation instruments interviewees knew about and what do they do think works well or does not work well with those instruments?
5. Interviewees opinions on quantitative vs. qualitative measures.
6. Other than a standard evaluation instruments, what do interviews believe The Center could do to gain insight into lessons learned?

Interviews were hand coded[2] by question and topic. Topical coding focused on:

- The 'most important' information a project evaluation instrument can produce
- Types of measures (or indicators) an instrument should include;
- How an instrument should be deployed;
- Other instruments currently in use;
- And other evaluative methods The Center could use. (The baseline evaluation instrument assumed focuses on surveys or questionnaires.)

Measures were topically into six categories.

1. Hard Measures
   These are agreements, durability of agreement and participant satisfaction. These three measures have been considered 'traditional' to some degree.
2. Soft Measures
   Soft measures are more numerous than hard measures and include reports on transparency of process, mediator skill, fairness of process, participant access to information, etc.
3. Soft Measures – Transformative
   Transformative measures capture how parties are changed due to the ADR process. This

[2] Interview coding is a process of categorizing and indexing interview responses.

includes changed levels of trust, the ability to work collaboratively in the future, and how people's understanding of both the conflict substance and the conflict process are altered.

4. Participant Perceptions
   Participant perceptions are a sub-category of all of the above and specifically highlight the reliance on measures as being reported by the participants. Most soft measures rely on participant perceptions being reported, however there is some amount of practitioner recognition that this deserves its own frame.
5. Quantitative and Qualitative methods of data collecting
6. Other

After reporting on the results of the coding this summary will touch on anecdotal feedback reviewed as well. Below is a chart showing interview results. A more complete narrative report is contained in the appendices. Note: interview questions were posed in an opened fashion: interviewees were not asked to choose between possible answers. For example, they were not presented with three options for the first question 'Most Important Thing…', their answers were voluntary. The categories of answer types below are a product of interview coding.

Most Important Thing to Learn From Project Evaluation	
Outcomes for agreements, durability of agreements, and participant satisfaction	13
Outcomes for process improvement and feedback loops	14
Other: process improvements, lessons learned, participant perceptions, comparisons to process BATNA, and mediator skill	31
Types of Measures or Indicators	
Hard Measures	
Agreement Reached	9
Durability of Agreement	6
Participant Satisfaction	10
Cost/Time	15
Misc	4
Soft Measures	
Participant Communication	11
Were the right parties involved?	7
Was the right information used?	6
Transformative measures	16
Participant Perceptions	24
Other: satisfaction with mediator, commitment or ownership, review by 'losing side', transparency, stakeholder needs, length of process, and procedural satisfaction	24
Qualitative vs. Quantitative Measures	
Both are useful	11
Quantitative only	2
Qualitative only	5
Other Instruments (respondents were aware of)	
No response	4
None	13

USIECR	10
EPA	1
"Frank Dukes"	1
Alternative Evaluation Methods (to surveys/questionnaires)	
Focus Groups	5
Stakeholder Interviews	5
Neutral Observer	5
Other: Long term follow up, internal staff meetings, score cards, regional forums, affiliations	11

The following questions were not coded.

How Should the Instrument be Deployed?

While not a direct question, topics such as how the instrument should be deployed, who should deploy it, and what shape it should take came up in half of the interviews. These responses assume a 'standard' instrument like a survey or questionnaire. The suggestions include pre-project, mid-project, and post-project questionnaires or surveys; deployed by the facilitator, a neutral 3rd party, or a project manager; issued by hand, via interview, or online (for example, Survey Monkey); and that participants as well the facilitator or mediator should provide some sort of report or feedback.

Other

There were ten uncoded responses. These were questions regarding the political and economic context surrounding the ADR project, whether or not participants would recommend ADR to a colleague or friend, how many parties dropped out, the level of leadership involved within the participant pool (essentially, whether or not participants have representational authority), and whether or not the project creates public benefit.

Anecdotal Feedback

The interviews produced a large amount of anecdotal commentary that was not coded, but was recorded. Much of this commentary was cautionary in nature and several interviewees expressed concern at the prospect of ADR evaluation for several reasons. Any measure of 'success' is bound to fail because ADR is essentially an iterative event within a larger process – the conflict and the history of events leading up to and away from the conflict. This essentially states that conflict resolution contains a great deal of uncertainty by its very nature. To measure an uncertain process with formulaic or static yardsticks will always result in abbreviated or misleading reports. In other words, it's very hard to use fixed and *certain* measurements with an uncertain process. We can think about this with the example of welfare to work programs. A person truly in need of welfare to work aid may have a very long term need that a 6 month or 1 year or 2 year performance measurement has no ability to capture. Because 'success' in ADR projects is so widely varied and exists separately in so many different opinions, many feel that evaluations that aim to impose an external framework for success have no possibility of achieving the evaluative aims. Simply put, we may not yet know what success means and more philosophically stated, we may not *be able* to know.

Another danger several interviewees mentioned was the potential misuse of evaluation, so that evaluation use must be very specific. For example, if ADR project evaluation states that a facilitator could have done a better job, it may jeopardize that facilitator's career. If that is to be the result of an evaluation then it *must* be one of the intended consequences of the evaluation. Further, what is the facilitator to be measured on? What is the criteria for facilitator evaluation? Some facilitators enter a project many years into the conflict resolution process and thus have influence in only a specific portion of the conflict.

However, on the whole, evaluation was seen as a lacking and deeply needed effort within ADR. ADR evaluation is complex, difficult, and requires delicate care and sophistication. The important take-aways from these interviews are:

- There are no definitive set(s) of questions, type or modes of questions, formats for implementation, or use of results which are universal. Because of this, an evaluation which relies on only one type of question or format is inherently weak. A strong evaluation will include multiple modes of questions (qualitative and quantitative) and types of questions (hard and soft).
- ADR project evaluation is an endeavor which fundamentally contains uncertainties. Those engaging in ADR evaluation need to remain cognizant of this fact and resist the desire to find certain and grounded universal ADR process facts.
- The purpose of the evaluation should be very clear: evaluation questions and implementation should reflect its purpose.
- ADR projects offer many types of value to participants and society. Defining 'success' in a narrow sense undermines the very benefits ADR projects offer.

Summary of Several Currently Used Evaluation Instruments

Evaluation instruments reviewed include those from the U.S. Institute of Environmental Conflict Resolution, the Environmental Protection Agency, Oregon Consensus, Oregon Solutions, Timothy Hedeen, PhD, Franklin Dukes, PhD, and others provided confidentially by practitioners. The instruments reviewed were all post-project evaluation surveys. As seen with the framework examples above, these instruments contained both similarities and differences. Most of the instruments combined quantitative with qualitative questions, while one was purely qualitative and one purely quantitative. Some were very long, others very short. Some emphasized questions about outcomes while others emphasized questions about process. The take-away from reviewing these instruments is that the field, as a whole, does not use generic evaluative principles or implementations, and that instruments differ depending upon their goals, origination, and type of project evaluated.

Words of Caution

The literature and the interviews offered many cautionary words of advice that are consistent with the program evaluation principles above.

With regard to all of the measures and varieties of possible implementations thus far touched on, there are cautionary principles that must be incorporated into developing and deploying an ADR evaluative process. Evaluation instruments must be tailored to the circumstances in which they are applied so that they yield credible and useful answers. For example, if one evaluation of ADR projects is used for both upstream and downstream processes (upstream disputes or projects occur early in the process prior to policy or regulation, while downstream projects entail issues which involve existing policy), it must at the very least detail the different natures of the project's respective positions lest the results be unusable. Or if one evaluation instrument used in very similar projects, and, all things being equal, is deployed at different intervals in different projects (shortly after consensus or perhaps nine months later) then the evaluator faces a problem of sample maturation and has created, *de facto*, two totally different types of evaluation. It could be that an evaluator would choose to do this in order to display changes over time, however that would itself be the purpose of that particular evaluative process.

Evaluations are essentially aimed at either creating credit or negating credit (either by way of outcomes or by way of process analysis – summative or formative evaluations). That being the case, evaluations, evaluation theory, and evaluation implementation must be done with specific and professional consideration lest there be potentially grievous unintended consequences. For example, an evaluation could be used to capture lessons learned and foster continual process improvements. Part of that evaluation might ask respondents to rate various aspects of mediator performance such as knowledge, fairness and the like, based on their perceptions. A disgruntled participant could use the evaluation as an opportunity to express frustration at part of the process that the mediator had no control over such as external political forces bearing weight on that participant. Subsequent publication of survey results could spell doom for the professional mediator who led the process, which was not the intent of the evaluation and does represent an influence outside of the process.

If an evaluation is a survey or questionnaire, the questions themselves must be crafted in a manner that does not unintentionally influence, or outright force, respondents into one answer or another. For example, if survey respondents are asked to answer a question on a non-likert cardinal scale, they are forced into either rejection or assertion with no 'neutral' option available (Patton 2008). Further, single-shot surveys may not capture how participant perspectives change over time. Additionally those surveys reply on people's memories (Conley and Moot 2003). Surveys need to be crafted to capture the correct participant group. Should a survey be restricted to the participants of the ADR process, or be extended to person affected by the ADR process outcomes?

With regard to soft transformative measures, Frank Dukes reports that at least some studies suggest that measures of relationship changes are sparse. The use of satisfaction-based criteria could lead to misuse of a meditative process. Some participants may use the evaluation as a means to create pressure on other parties and consequently report their high level of satisfaction (Dukes 2004). Foley reports that traditional measures of success, Agreement Reached, Stability of Agreement, and Party Satisfaction, have been said to rest on questionable assumptions regarding bounds of power, resources and cultural constraints. As a result of this, more relational measures have been developed. (Foley, 2004). Evaluations that target goals may not assess the appropriateness of the goals and objectives themselves or the assumptions they rest on. Further, goal based evaluation used in collaborative processes force those processes to have clearly defined goals that such processes do not always possess (Conley and Moot 2003).

To be a useful and effective management and planning tool a survey measurement device must provide a flexible process for reevaluating the goals of the program, modifying the evaluation methodology, and implementing necessary changes (Fairman 1999). Further, evaluations that are aimed a process improvements must be useable by those who design future processes.

In general, there are issues that evaluations should be cognizant of. Reliability concerns the extent to which a measure produces the same results when used repeatedly to measure the same thing. Validity is the extent to which an instrument measures what it is intended to measure. Sensitivity is the extent to which the instrument is capable of capturing subtle differences in what is being measured (measuring grams is more sensitive than measuring pounds).

“We should approach the research presented here and elsewhere as we should approach our own experiences as practitioners, researchers, program managers, funders, or ECR consumers: with an inquiring mind, a tolerance for ambiguity and uncertainty, and an expectation that we will always be hungry for more answers and understanding that allow us to do better work” (Dukes 2004).

TRADEOFFS

In order to provide The Center with critical and constructive recommendations, it is important to look closely at the benefits and costs and tradeoffs of different types of data collection (what sort of collection instrument: surveys, focus groups, etc), how data collection is implemented, and the types of questions respondents are asked to report on.

Evaluation Instrument Type

- Surveys / Questionnaires

Surveys or questionnaires can be done by participants with anonymity and can be implemented to allow respondents to fill them out at their own leisure. The greatest attraction to surveys is that they can be managed at relatively low cost to The Center, implemented easily, and can be constructed to allow easy analysis. However, surveys present a selection bias in that respondents may only provide feedback which is asked for and frequently have varied rates of response. Because of the fluctuation in response rates, The Center may not gather all the information it seeks. Participants may feel more or less comfortable responding on a 'successful' or 'unsuccessful' project. The Center would have to be diligent in collecting survey responses and potentially request feedback more than once which could disaffect project participants.

Surveys can be implemented at different points: Post project; pre and post project; and during some mid-point through a project in addition to post and/or pre/post. Frequently the types of questions asked in project evaluation ask for participant reflections or perceptions on topical areas. Questions of these sorts, while probably necessary, are hindered by the fact that each respondent has unique experiences which inserts a relative nature into the survey responses. A pre/post survey structure can allow The Center to legitimize some of the post survey responses. A pre survey constructed to mirror some of the material gathered in the post survey would allow a one to one before and after comparison with each participant, thus allowing a more detailed look at the effects of the project on each individual participant. Drawbacks of a pre/post format are that they are more costly and resource intensive than a post survey only format and they may introduce a sort of Hawthorne effect into the process. However, it can be argued that this Hawthorne can actually benefit both the process and the participants in that it essentially coaches participants. This is important because participant training has been identified as an important part of ADR processes.

A mid-project survey evaluation can potentially allow the facilitator or mediator to gain important insights into how participants are feeling or thinking and thus allow important issues to come to light or allow for process course corrections. However, pausing what is essentially an intervention in order to perform a higher level intervention can seem confusing and introduce kinks into what is a dynamic and fluid process (ADR). Further, a formal pause in an ADR process for an evaluation to take place can be avoided through backchannel communications between the facilitator and process participants.

Surveys can be implemented in a number of formats. Participants can receive surveys via handouts in person, via email, use a service such as Survey Monkey, or respond to them through semi-structured interviews. Emailed documents, Survey Monkey formats, and to a lesser extent handouts, allow respondents to complete in their own timeframe, as indicated above, provide some degree of anonymity of chosen (which for many people may generate a greater willingness to be forthcoming), and require relatively low cost to The Center. A benefit of a Survey Monkey style service is that analytical tools are part of the service's features. Again, the data collected in these manners can be limited to what is on the screen or piece of paper. Semi-structured interviews are high cost to The Center and provide no anonymity to participants. However, the ability for deeper dialogue, and thus data gathered, is pronounced. Further, one on one feedback and attention lets participants know that their thoughts and feelings are truly valued which invests them in the process and thus creates, perhaps, more reliable information. Semi-structured interviews can create strong relationships and help create goodwill toward ADR as a practice.

Many participants may not wish to participate in one or the other which creates a sampling issue. Because of this it may be useful to use a hybrid model where, for example, a Survey Monkey survey would be distributed to project participants wherein they would be asked if they were willing to meet in person or via the phone for a more detailed conversation. This hybrid model mitigates cost to The Center (in terms of interview hours) and allows participants to choose their feedback method. While respondents would self-select to each feedback method, this will still allow The Center to gather more data. However, it presents a problem for aggregate data analysis in that there would be two different evaluative instruments. This may not present a huge issue for The Center because the purpose of this evaluation instrument is to capture lessons learned and foster process improvements. If The Center were to embark on larger scale empirical research, there would need to be some compensation for the difference in evaluation instruments. At the very least the data collected from each would need to be presented as such.

- Open ended participant interviews

These are similar to semi-structured interviews based on survey formats mentioned above. Their difference is that they are intentionally opened-ended to allow participants to engage in a conversation about the process they participated in rather than asking them to answer a specific set of questions. This presents analytical problems when looking at aggregate data, transforms feedback into anecdotal information rather than standardized, and comes at high cost to The Center. Because of the time cost and intimate nature of this method, participants may have mixed feelings which could result in feedback from only those who self-select. Open-ended interviews may provide a great deal of sensitivity, though the broad reliability of them is questionable. (However, demonstrating reliability could potentially be achieved through detailed interview coding.) The transactional cost and selection bias of interviews and their coding can be higher than with other methods.

- Focus Groups

Focus groups are an alternative to variations of standard project evaluation. Focus groups ask participants to reconvene at some point not too distant after a processes conclusion to discuss various aspects of the process (just as surveys ask participants to give feedback on various aspects). The costs are somewhat high in terms of time and resources to both participants and The Center in that a focus group is another meeting which requires preparation, moderation or facilitation, and summarizing. There are selection issues as not all process participants will be available to reconvene for a focus group, and asking for that commitment up front (perhaps in ground rules) is itself a barrier to entry in the ADR process. Further, in some projects it would be impossible to ask all participants to reconvene, The Center faces the painful task of choosing who to invite. However, a focus group can provide a type of feedback which is otherwise unattainable because participants in the focus group build off each other's comments and feelings, and through discussion are able to discover details of the ADR process which they may not have been able to by filling out a survey individually. If an ADR process is about a group of people finding a solution to a problem which could not be discovered by individuals, then a focus group is the mirror of this within the bounds of evaluation. On the downside, focus groups may also face challenges in terms of reliability and validity.

- Observation based

A third party neutral observer, acting as a process evaluator, could silently participate in a Center project and report on their findings. This method could possibly provide extremely detailed feedback, but would come at very high financial cost to The Center. Further, it would limit feedback to The Center to a single point of view. A fuller account would require observer based evaluation in combination with a second method of feedback.

- Case Studies

Another method for creating process improvements and identifying lessons learned is through the use of written case studies. A series of detailed case studies of Center projects could provide useful feedback and information for the development of future Center efforts. Case studies are an arduous task and come at a high cost to The Center. Alternatively, The Center could employ graduate level interns to write case studies, a resource which is at The Center's disposal because of The Center's relationship with UW and WSU. Case studies are particularly useful for assessing impacts and processes of a program (or project) which is unique, or one in which impacts are impractical or difficult to measure. Because of that they can be well suited to the sorts of projects which The Center engages in. Further, traditional evaluation techniques presuppose that the evaluated process assumes some rational and predictable path. Because ADR processes often take unexpected turns, the case study can be a valid way of capturing what happened, why, and the outcome in a way which traditional evaluation struggles with.

- Regional Practitioner Forum

Another method for capturing lessons learned and fostering process improvements The Center could engage in is collaborating with other practitioners. To this end The Center could host a regional mediator/facilitator forum in which practitioners from the Pacific Northwest (and perhaps nationally) could meet for several days, have panels and discussion groups and learn from each other's experience. The downside to this, clearly, is that it ultimately has no direct bearing on the projects which The Center directly engages in, either in the present or the future. It is certainly a wonderful learning tool for both The Center and the field. However, it does not provide direct project or programmatic feedback to The Center, and thus needs to be seen as a separate order of learning.

Evaluation Participants - Sampling

While it is clear that a project evaluation of this sort typically relies on information collected from project participants, with either surveys or interviews or focus groups, there are other sources of feedback as well. Project dropouts can offer an important perspective on project processes and outcomes, though gathering information from dropouts can pose complexities and require large amounts of time and effort. Community members impacted by the project but not part of the decision making process can also offer an important point of view for The Center to learn from, but again it can come at increased costs in terms of time and effort. Further, projects which The Center assesses that contain parties who choose not to engage in ADR comprise potential sample. While this does not directly relate to The Center's goal of *process feedback*, it is nevertheless a potentially useful feedback source for The Center. Who participates in project evaluation will be a sensitive issue regardless of what type of evaluation The Center uses or how it is implemented.

Instrument Facilitator

The Center's evaluation instrument (whether survey, interview or focus group) will need to be implemented by someone. Candidates for instrument implementation are: the project's facilitator or mediator, a Center staff member related to the project (or perhaps graduate student intern), or a neutral third. Benefits of the evaluation being administered by the project's facilitator or mediator are that they would have a deep understanding of the project in question and potentially positive relationships with project participants. On the other hand, process participants who have less than stellar feelings about either the process or the practitioner may be less inclined to interact with the practitioner or provide feedback to them. It is perhaps easier to sing praises than it is to condemn. Another issue with practitioner administered evaluation is unintentional bias being introduced into the evaluative process. Further, if feedback is requested from participants regarding the practitioner, practitioner administered evaluations face an awkwardness which could result in compromised data collection.

A neutral third administering the evaluation would not present some of the challenges with practitioner administered, however a neutral third also would not be able to benefit from relationships built between practitioner and participant. Additionally, a neutral third would probably not have expertise in the topic are. While this would not interfere with the administration of surveys, it would pose a hurdle for feedback collected through an interview process. The main attraction to a neutral third comes built into their designation: neutrality.

Another possible candidate for project evaluation administration is a project manager. Someone related to the project and familiar with the participants and the process history, yet someone distant from the project itself. This could present a compromise between the two previous options.

All of three of these options and their considerations bear more weight on evaluation strategies which use one on one communication. For a purely survey method of collection the issue of how administers becomes only an internal issue. That internal issue is mainly concerned with time and cost and the possibility of bias. In either case, it seems that having the practitioner administer the evaluative process presents a possible conflict of interest for The Center.

Implementation Timeframe

Another issue to consider is when the evaluation takes place, regardless of its format. An evaluation which takes place directly after the end of a process may have its results skewed by 'post process glow/unhappiness'. This runs the risk of capturing the feeling of a particular moment of a process rather than capturing feedback about all phases of a project. On the other hand, evaluation delayed for a long period of time after the completion of a Center project runs the risk of diminished participant recollection.

Instrument Size

Regardless of the method used for evaluation, it must be useable by those involved. For surveys, those which are very short (under 10 questions) are easier for participants to complete while sacrificing the amount of data collected. Surveys which are very long (30+ questions for example) or complex can provide more information as the cost of potentially inhibiting respondents from completing.

Survey Question Types

Along with balancing instrument size is the challenge of collecting enough of the *right* information so that the evaluation is able to meet its goals. An evaluation instrument which employs only hard measures may be easy for participants to engage in and be somewhat easy for analysis, however it would lack a certain depth necessary to gain reliable insight into a project's process. Hard measures on the surface can appear reliable and valid, but because of the differences in participants themselves, hard measures may be somewhat superficial. Reports on soft and transformative measures can provide information on the process and its impacts on participant experiences. For gaining insight into lessons learned from Center projects this can be useful, however again there is the problem of reporting being relative to different perspectives among project participants. This issue is of course present with evaluation reporting which relies on participants reporting their perceptions as well.

Close-ended questions are easy to analyze and enhance reliability but provide less depth and substance and may inadvertently force invalid responses. Open-ended questions are more difficult to analyze and may require coding. They also present the difficulty of understand a respondents *meaning* via their words. However, open-ended questions allow for more detailed responses and can provide insight which close-ended cannot.

Likert scale questions are easy to answer and analyze, however they lack the depth and nuance necessary to identify lessons leaned and foster process improvements. Qualitative essay style feedback may present difficulty in analysis, require considerable coding, and strain respondents. Using them in combination mitigates some of the issue.

Reliable and useful surveys should present questions in a straight forward manner, free of complications such as double barrel questions, double negatives, etc. The questions themselves should be written in a manner which does not create uncertainty and propel answers in one direction or another, and they should seek to create exclusive and exhaustive categories.

RECOMMENDATIONS

My research began with the expectation of identifying weak evaluative methods and excluding them, in order arriving at something 'really good'. However, the examination has led me to the conclusion that there are many evaluative techniques that would produce useable results for both The Center and the field in which The Center operates.

While formulating recommendations the following criteria were used:

- How options rate against the program evaluation principles and biases
- Transactional cost
- Usefulness to The Center

(See Appendix II for a visual matrix of tradeoffs)

Main Recommendations

Considering the options, suggestions and information gained from the literature, interviews, existing surveys, and tradeoffs (above) in order to reach The Center's goals for this evaluation instrument (capture lessons learned, foster best practices and continually create process improvements), using the criteria above suggests that The Center:

1. Utilize a pre/post evaluation design in the form of surveys or questionnaires.
    a. One of the drawbacks of a post process only evaluation relying on participant feedback is that such feedback is often personal. Respondents may interpret questions slightly differently. Using a pre-process evaluation can help in two ways. First, it can be used to help passively coach participants in the values inherent in collaborative processes. Second, it can set a baseline of information for each respondent, which can be analyzed side by side with their post process survey responses. Because of the difficulty obtaining a comparison group, a pre/post design will help The Center draw inferences and attributions.

2. Gather feedback from both process participants and process practitioners.
    a. This is for two reasons. First, the practitioner questionnaire can serve a project management role by gathering project information such as number of meetings, hours spent in meetings, hours spent working on the project outside of meetings, and so forth. Second, practitioner feedback on the process and outcomes can be viewed alongside participant feedback, which helps create a robust view of the project by contrasting opinions from all sides of the table.

3. Focus evaluation surveys on four key elements:
    a. Process Substance
        i. Right questions, right people, right information, etc

b. Process Procedure
    i. All participants had equal voice, information was informative, meeting structure supported dialogue, etc
c. Process Transformation
    i. Trust gained, relationships altered, learning of key interests was broadened, etc
d. Process Outcomes
    i. Agreements reached, participant approval, implementation prospects, etc

These four elements capture several things. First, they loosely reflect a logic model framework. This is important because it allows easy conceptual interaction and analysis of the instrument, its results, and any adjustments made to it (which would make it a new instrument and cautions therein would be heeded). Second, it captures multiple types of outcomes. This is important because for any process in which The Center engages, there are multiple types of success that can be reached. Restricting evaluation to only certain types of success, or avoiding the issue entirely, creates less useful feedback from which The Center can learn.

4. Implement evaluation no later than three months after the final meeting of the project. Ideally, there would be a short delay after the final meeting to allow emotions to cool and outcomes to be apparent, but this may not always be logistically feasible.

5. Keep survey length to between 15 and 30 questions. Shorter would provide too little information, longer presents an obstacle to completion.

6. Remain flexible on whether the survey is administered by handouts or by an online tool such as Survey Monkey, and use either or both as needed and informed by The Center's experience.

7. Remain open to utilizing both survey/questionnaire feedback methods as well as in person semi-structured interviews. The former allows participants to respond on their own time and gives them a degree of autonomy, while the latter may be useful in creating more detailed responses as well as continue to foster positive relationships.
    a. In person semi-structured interviews should be administered by Ruckelshaus Center staff other than the practitioner of the project, to avoid conflicts of interest and create a potentially safer space for respondents. However, this is not a hard rule. There will be projects where the relationship between practitioner and participants is very strong and may increase participation in semi-structured interviews. The Ruckelshaus Center will have to exercise judgment in this matter.
    b. In the case of semi-structured interviews, the Ruckelshaus Center should begin with the survey, but ask participants if they would be willing to meet in person or via phone for the semi-structured interview.

   c. The Ruckelshaus Center should analyze the results of interviews and surveys as complimentary to each other, not in conjunction with each other as they are different instrument types and combined analysis would be unreliable. However, semi-structured interview questions should categorically mirror survey questions.

Secondary Recommendations
The Ruckelshaus Center may wish to consider the following, allowing for time and effort.

1. Survey Community Members
   a. Ruckelshaus Center projects impact not only the participants of those projects, but the community at large. The Ruckelshaus Center could seek input on how the community affected by a project feels about that project's outcomes and how the community feels about the processes which the Ruckelshaus Center uses. The types of projects the Ruckelshaus Center engages in are frequently transformative and relationship building. The Center could seek to foster good relationships with the community as part of its evaluative efforts.

2. Follow up on Project Agreements
   a. The Ruckelshaus Center could monitor the implementation of agreements created as the result of its projects. Documentation of ongoing implementation of project agreements is in The Center's interest. Frequently and method would be best decided by Center staff, however a minimum of annual progress report is reasonable.

3. Focus Groups
   a. When possible and where the Ruckelshaus Center deems appropriate, The Center could reconvene project participants in a focus group in order to create sustained and dynamic discussion and reflection on the project. This would not be a quantifiable effort in the same manner as surveys. It would be purely qualitative and should be seen as such. Gathering participants in an open and warm environment and simply letting them talk about the project and process can deliver important insights into not only Center process, but Center clients and the surrounding community.

4. Case Studies
   a. The Ruckelshaus Center could consider writing case studies of its projects. This serves as another lens through which to view projects and thus creates a new venue from which to learn. Case studies are frequently written by graduate students. This aligns with The Center's goal of advancing the teaching and research missions of the two universities. Lastly, case studies can be distributed to other practitioners and centers in the field, adding value to the field's total knowledge.

Options not recommended:

a) Do not employ neutral observers. Observers are very expensive in terms of time and cost, but more importantly because their potential for disaffecting project participants.

b) Do not gather extensive feedback on practitioner quality. That information can be gathered just as effectively by other means while not creating the potential professional and political complications such feedback creates.

c) Avoid mid-point process interruptions for mid-point evaluations. Mid-point evaluations run the risk of damaging an in-process project. If the goal of a mid-point evaluation is to find the temperature of the group, make course corrections, identify loose ends, etc, the practitioner can accomplish the same ends with one on one participant check-ins, back channel communications, and individual meeting exist surveys.

d) Avoid take-home questionnaires. Take home questionnaires have a low response rate which creates both sample issues and data issues for The Center.

INSTRUMENTS

1. Post Process Participant Evaluation
   This questionnaire style survey combines elements of the USIECR Facilitation and Mediation Participant Surveys, surveys used by Frank Dukes, and is otherwise informed by the literature review and interview process.

2. Post Process Evaluation Questionnaire for Semi-structured Interviews
   This questionnaire uses the same questions as the above but is altered for semi-structured interviewing.

3. Pre Process Participant Questionnaire
   This questionnaire is based solely on the feedback from interviews and is designed to have some aspects parallel to the Post Process Participant Evaluation

4. Post Process Practitioner Survey
   The post process practitioner survey is a nearly identical with the USIECR Facilitation Practitioner Questionnaire

Post Process Evaluation Questionnaire

The William D. Ruckelshaus Center performs evaluation of its services, projects and processes. Your responses to this questionnaire will help The Center identify lessons learned and create process improvements so that we may better serve our communities and citizens. Responses to this questionnaire are confidential: key themes and findings are shared, but identity of individual respondents is not recorded unless offered by the respondent.

This questionnaire should take about 20 minutes to complete.

Please indicate what organization or interest best describes you (Federal government, Tribal Nation, concerned citizen, etc) *[drop down box in electronic version]*

Please identify the number which best indicates your level of agreement:
1 = strongly disagree, 2 = disagree, 3 = neutral, 4 = agree, 5 = strongly agree

Process	Disagree ---- Agree				
The views and interests of all participants were considered during the process	1	2	3	4	5
All participants in the process were treated respectfully	1	2	3	4	5
All participants were able to participate in the process	1	2	3	4	5
You were able to participate in the process	1	2	3	4	5
The process was not dominated by one or more person(s)	1	2	3	4	5
All the people or interests needed to participate in this process were included	1	2	3	4	5
If not, who do you feel should have been involved who was not?					
The end goals of the process were widely agreed-upon and clearly articulated at the outset.	1	2	3	4	5
If goals changed along the way, that was accomplished and articulated in an intentional and open manner. N/A	1	2	3	4	5
Interim goals, milestones or benchmarks were clear N/A	1	2	3	4	5
Activities (meetings, breakout groups, etc) the process engaged in were useful in helping the group move forward	1	2	3	4	5
You, at any point, were positively 'coached' on how to participate in a collaborative effort.	1	2	3	4	5
Process activities proceeded in a fair and unbiased manner.	1	2	3	4	5
Please provide any additional feedback on these questions below.					

Information	Disagree ---- Agree				
The *right* information was used during the process.	1	2	3	4	5
That information was accessible and understandable to all participants.	1	2	3	4	5
That information was accessible and understandable to you.	1	2	3	4	5

The information used in the process helped the group move forward.	1	2	3	4	5
This process identified *all* relevant and necessary issues needed to make progress	1	2	3	4	5
Please provide any additional feedback on these questions below.					
Relationships	Disagree ---- Agree				
The process improved communication among all parties	1	2	3	4	5
The process improved communication among 'key' parties	1	2	3	4	5
Trust was built between participants as a result of this process	1	2	3	4	5
This process helped you to understand other's interests and values	1	2	3	4	5
Participating in this process will help you in future collaborative efforts	1	2	3	4	5
Please provide any additional feedback on these questions below.					
Outcomes	Please select				
The project reached agreements on:					
All key issues	☐				
Most key issues	☐				
Some key issues	☐				
No agreement was reached but progress was made toward resolution	☐				
No agreement was reached or progress made	☐				
	Disagree ---- Agree				
The project's outcomes reflect all perspectives brought up during the process	1	2	3	4	5
The project's outcomes will solve the intended problem(s)	1	2	3	4	5
The goals stated at the project's outset were met	1	2	3	4	5
Goals which may have arisen during the process (if any) were met. N/A	1	2	3	4	5
The projects outcomes, if they include agreements, can be implemented.	1	2	3	4	5
If the project's outcomes have 'next steps' for you, you fully understand them.	1	2	3	4	5
If so, you are committed to the 'next steps.'	1	2	3	4	5
In general, you feel this was a good process to solve the problem.	1	2	3	4	5
Please provide any additional feedback on these questions below.					

What do you feel were the best and worst parts of this process?

What surprised you about this process?

Do you feel that the process resulted in any negative impacts?

If you could have changed one or more parts of the process, what would it/they have been?

Please take a moment to elaborate on any of the questions asked in this questionnaire, or address any other topic of your choice.

Would you like to be contacted for a follow-up conversation, to provide a more in-depth evaluation of this project? [yes/no] If yes, please provide contact name, email and phone number. Note, The Center may contact you proactively to request such a conversation.

Post Process Evaluation Questionnaire for Semi-structured Interviews

Thank you for taking the time to speak with me and your willingness to share your feelings and insight regarding the [name] project. The William D. Ruckelshaus Center is interested in evaluating its projects in order to capture lessons learned and foster process improvements. The questions I have today are similar to the survey sent to you previously, but are more open ended to allow us to have a fuller conversation. This interview should take about 30 to 45 minutes depending upon how much information you would like to share. Information collected from you today is confidential. Are you ready to begin? (Interviewer: sub-questions are for prompting dialogue, use as needed)

1. What is your affiliation with regard to this project? Who or what did you represent?

2. During this project, do you feel that all the interests of all the participants were heard?
   a. Was there an incident you can remember where participant interests were not voiced or were voiced, but not included in the discussion?
   b. Were your interests taken into consideration during the process?

3. Do you feel the process was fair to everyone who participated? If not, please explain.
   a. Was the processes dominated by one or more voice or personality, in totality or at any particular phase?

4. Do you feel the goals of this process were clearly and articulately stated at the process outset? Did the group play a role in developing and/or confirming those goals?
   a. How did you feel about those goals?
   b. Did they change or evolve during the process? If so, was that an intentional or transparent evolution?

5. How did you feel about the activities (meetings, field trips, information sessions, etc) that happened during the project?
   a. Do you feel that any were redundant or otherwise lacking?

6. When thinking about the information this processed used, do you feel that it was adequate?
   a. Did the information help the group move forward?
   b. Was there information you wish had been used which was not?

7. Do you feel that the relationships between participants changed as a result of this process?
   a. If so, how?
   b. Do you feel that your own relationships changed?
   c. Has your opinion of other stakeholders changed?

8. Did the outcomes of this process meet your needs?
   a. If not, please explain.
   b. Did the process meet other participant needs?
   c. Would you call these outcomes successful for all involved?

9. Do you feel that The Center being a university-affiliated practitioner had an effect on the process (positive, negative or both)? In what ways?

10. What do you feel were the best and worst parts of this process?

11. If you could have changed one or more parts of the process, what would it/they have been?

12. Do you feel that the process resulted in any negative impacts?

13. Is there anything else you would like to give us feedback on?

14. Can you think of a question which I should have asked you but did not?

Pre Process Participant Questionnaire

The William D. Ruckelshaus Center performs evaluation of its services, projects and processes. Your responses to this questionnaire will help The Center identify lessons learned and create process improvements so that we may better serve our communities and citizens. Responses to this questionnaire are confidential: key themes and findings are shared, but identity of individual respondents is not recorded unless offered by the respondent.

This questionnaire should take about 20 minutes to complete.

Please indicate what organization or interest best describes you (Federal government, Tribal Nation, concerned citizen, etc) *[drop down box in electronic version]*

Please indicate how you feel about the following with
1 being negative, 3 being neutral, and 5 being positive.

At this point:	Very low- Very High					
How well informed do you feel about the issues involved in this policy challenge?	N/A	1	2	3	4	5
Your overall knowledge of other parties' interests and concerns?	N/A	1	2	3	4	5
Your overall level of trust for the other parties involved?	N/A	1	2	3	4	5
If you'd like to indicate your level of trust for particular parties, please do so here:	N/A	1	2	3	4	5
	N/A	1	2	3	4	5
	N/A	1	2	3	4	5
Your overall amount of communication with the other parties?	N/A	1	2	3	4	5
The overall quality of your communication with other parties?	N/A	1	2	3	4	5
(Please provide any comments about amount or quality of communication with specific parties here)						
Rate your level of understanding of [type of process] processes	N/A	1	2	3	4	5
Rate your level of confidence that this process is the right one to address this/these problem(s)	N/A	1	2	3	4	5

Please identify the number which best indicates your level of agreement:
1 = Strongly Disagree, 2 = Disagree, 3 = Neutral, 4 = Agree, 5 = Strongly Agree

I expect this process to:	Strongly Disagree – Strongly Agree					
Include all the necessary parties needed to have a full discussion of the problem	N/A	1	2	3	4	5
Identify all the relevant issues involved in this problem	N/A	1	2	3	4	5
Alter relationships between the involved parties	In a positive fashion					
	In a negative fashion					
	In no meaningful way					
	I have no guess					

The project will reach agreements on:						
All key issues	☐					
Most key issues	☐					
Some key issues	☐					
No agreement will be reached but progress was made toward resolution	☐					
No agreement will be reached or progress made	☐					

Please take a moment to write down your knowledge of, or experience with, the William D. Ruckelshaus Center

Do you feel clear about the goal(s) of this project? Can you briefly state them?

Please take a moment to elaborate on any of the questions above or on any of your expectations for this project

Thank you very much for your time and participation!

Post Process Practitioner Survey

The William D. Ruckelshaus Center performs evaluations of its projects. Recently, you have served as a practitioners in one of these projects/cases, and the Center requests your assistance with this project evaluation. The purpose of the Center's evaluation is to identify lessons learned, identify best practices, and foster continual process improvements. Your responses will provide information that will be used to improve our programs and services. If you have any questions or concerns please contact *email address*.

1. What was the application of the collaborative process in this case or project?
OUT OF THE FOLLOWING LIST, PLEASE CHECK THE MOST APPROPRIATE BOX.

Please select those that apply	
Policy development	
Planning	
Siting and construction	
Rulemaking	
License and permit issuance	
Compliance and enforcement action	
Implementation/monitoring agreements	
Other (please specify): ___________________________	

Please identify the geographic location(s) for this case:

City(s), County(s), Watershed(s), etc.________________________________

State or States ____

Regional ____

Please identify the central issues to this collaborative process: (Check all that apply)

	Agriculture
	Air Quality
	Archeology or Historic Preservation
	Coastal Zone or Marine Management
	Collaboration Training
	Education
	Economic Development
	Ecosystem Management
	Endangered Species and/or Critical Habitat

	Energy
	Forest and Timber Management
	Governance
	Land Use and Urban Development
	Mining
	Native American, Alaska Native, Native Hawaiian Issues
	Parks and Refuges
	Public Health
	Recreational Use and Access
	Solid or Hazardous Waste
	Transportation
	Watershed/River Basin Management
	Water Quality
	Water Supply
	Wildlife Management
	Other

Approximately how many sessions (including conference calls and meetings) were held in conjunction with this particular process?

______________	Number of sessions

Estimate the number of hours you and any other practitioners devoted to this case/project:

______________	Total hours

What was the total number of months or years in which you were actively working on this case/project:

0 – 6 months	
7 – 12 months	
1 – 2 years	
3 – 5 years	
More than 5 years	

On a scale of 1 to 10, with 1 being 'not at all difficult', 5 being 'moderately difficult' and 10 being 'extremely difficult', please rate the difficulty of:

Developing an effective collaborative process for this case or project ____

Implementing an effective collaborative process for this case or project ____

Did The Center provide you with the resources, time, and information you required to hold an effective process?

____

____

If not, what other resources, time or information do you feel you (or the project) would have benefited from?

____

____

____

____

What was the greatest challenge that YOU faced as the practitioner in conducting an effective collaborative process (including barriers to participants' ability to participate)?

____

____

____

____

____

Please characterize the participants by doing the following:
Please indicate THE NUMBER of participants that were included in the process within each of the given categories *(e.g., if 2 separate participants represented the state, place a '2' in the box for 'State Government').*

Number of participants included in the process	
Academic	______
ADR Practitioners	______
Agriculture	______
Aviation/Aerospace	______
Consulting	______
Education	______
Environmental Interest	
Federal Government	______
Finance	______
Forest Products	______
General Public	______
Healthcare	______

Information Technology	_______
Labor	_______
Legal	_______
Local Government	_______
Marine Industry	_______
Media	_______
NGO	_______
Philanthropy	_______
Real Estate	_______
Retail	_______
State Government	_______
Tribal	_______
Special Advocacy Interests (Please specify): ________________________________	_______
Other (Please specify): ________________________________	_______

On a rating scale of 1 – 10, with 1 being Strongly Disagree and 10 being Completely Agree, please rate the following questions (Circle N/A if not applicable):

On reflection, I feel I was the right practitioner to guide this process.
Rating _________ N/A

If needed, resources were available to obtain the relevant expertise/information for this case or project
Rating _________ N/A

Experts were used to educate participants in the collaborative process on the relevant issues
Rating _________ N/A

In general, the relevant information was understood by the participants
Rating _________ N/A

Participants worked to ensure agreement on the meaning of the relevant information
Rating _________ N/A

On a rating scale of 1 – 10, with 1 being Strongly Disagree and 10 being Completely Agree, please rate the following questions (Circle N/A if not applicable):

Participants had the skills required for participating effectively in the collaborative process
Rating _________ N/A

Participants had the time required for participating effectively in the collaborative process
Rating _________ N/A

Participants had the financial resources required for participating effectively in the collaborative process
Rating _________ N/A

Participants had access to the information required for participating effectively in the collaborative process
Rating _________ N/A

Organizations or interests that should have been included in the process did in fact participate in the
Rating _________ N/A

Please indicate the extent to which progress was made:

(Check only one)

Progress made on all key issues	Progress made on most key issues
Progress made on some key issues	No progress was made

Use the space below if you would like to elaborate on your response:

Are there any relevant lessons learned that should be recorded? (Were there any specific events or issues that affected process outcomes? What was the 'most useful' aspect of the process? The most damaging? etc)

Would you like to be contacted for a follow-up conversation, to provide a more in-depth evaluation of this project? [yes/no] If yes, please provide contact name, email and phone number. Note, The Center may contact you proactively to request such a conversation.

Thank you very much for your time and input!

Considerations for instrument usage – Post project evaluation questionnaire

The Ruckelshaus Center should consider the following suggestions for using instrument results over time.

- Identify which questions or groups of questions lead to actionable information.
- Identify which questions, over a number of instrument usages, contain the highest amount of divergence in responses.
- Identify questions, if any, which show no difference in response over a ‘large’ number of uses. These questions may need to be refined to contain a greater degree of sensitivity, or perhaps removed.
- Identify which *projects* The Center is most *satisfied* with; analyses the instrument feedback for just those projects, then compare with instrument feedback from projects The Center is most *unsatisfied* with.
- Use feedback from project drop-outs and participant interviews to identify ‘missing questions’.
- Look for correlation between project statistics (duration of process, number of participants, number of meetings, etc) and survey data.

APPENDICES

Appendix A
Narrative Summary of Semi-structured Interview

From interview responses:

Most Important thing to learn

Out of 58 comments 27 of them focused on understand outcomes. Among those there was an equal amount of feedback directed at solidly measurable outcomes – agreements, durability, and satisfaction, as there was to using outcomes to create a feedback loop into process improvements. A small minority of these 27 indicated that both are equally important and inter-related. Several caveats were offered that specified that the most important thing an evaluation can capture is directed by the intent of the evaluation's implementation. Other responses included comments directly related to process improvement and lessons learned, understanding participant perceptions, gathering real time feedback, understanding participant BATNA to ADR, and practice of the facilitator or mediator.

Measures

Reponses on what measures an evaluation should seek to capture was easily the most diverse and complex set of responses. Answers interview respondents gave regarding measured was gathered in two different questions (for a total of 54 responses). Please note that respondents were not limited to a single response per question, though segmented responses rarely exceeded 4. The first asked their opinion on what sort of indicators would indicate success or danger for an project of the nature The Center engages in, and what specific questions they would want to see in an evaluation with the specific purpose of creating process improvements and capturing lessons learned.

Hard Measures

There were 40 comments which focused directly on hard measures. These include agreements, durability of agreements including longitudinal effects, participant satisfaction, and cost and time. Of the these 40, nine were about the agreement itself being reached, six regarded the durability of the agreement (which implicitly means agreement was reached), ten touched on how satisfied with outcomes participants felt, eleven were cost and time, there were four uncoded answers. These were: rate of agreement, whether or not something of value which is enduring was created, whether a high level of benefit was gained, and lessons learned for process improvements.

Soft Measures

There were 48 answers coded as soft measures. Eleven of these targeted participation and stakeholder communication, seven asked if all of the right parties were part of the process, and six asked if participants were well prepared and had access to the right information. These were the highest categorical responses. Others included satisfaction with the mediator / facilitator, commitment or ownership, a review by the 'losing side', process transparency, stakeholder needs, length and density of the process, procedural satisfaction, and capturing participant alternatives to ADR.

Transformative Measures

There were 16 comments that touched on 'transformative measures'. These were how much trust was gained between parties, how much trust was gained between mediator and parties, how relationships changed during the process, how parties relationship to the area of conflict changed as a result of the process, the participant's ability to work collaboratively in the future, development of negotiation skills, and participant's emotional satisfaction.

Participant Perceptions

Participant Perceptions are measures which were frequently covered in Hard, Soft, and Transformative measures, but the interviewee specifically referred to the perception of the stakeholder or participant. These also included perceptions of fairness, efficacy, and appropriateness of the process. There were 24 responses which referred to a participant's perception.

Quantitative and Qualitative Measures

Interviewees were asked their feelings on quantitative vs. qualitative measures.

Five were in favor of qualitative measure and said that quantitative measures should not be used. Two reported the exact opposite. Two reported positively in favor of quantitative measures while not commenting on qualitative measures. Eleven reported in favor of both types of measures, and seven did not comment on the question or reported 'I don't know'.

Other Instruments

Interviewees were asked if the knew about or used other ADR process/project evaluation instruments. Ten said they knew about the work done by USIECR, of which four currently use variations of, one referred to the EPA evaluation instrument (which is essentially the USEICR instrument), one referred me Frank Dukes' work, two did not answer the question and 13 reported no. Of those, the work by Juliana Birkoff and Peter Adler was mentioned. There was a general lack of satisfaction with the instruments currently available or the lack thereof and some amount of response which was pointedly negative about the prospect of evaluation itself (more on this later).

Other Types of Feedback

Interviewees were asked what other methods of process feedback The Center could utilize. The three most common responses, five each, were Focus Groups, post project Stakeholder Interviews, and using neutral observers. Other responses included engaging in long term follow up of project outcomes, internal staff meetings or 'mediator' meetings, writing case studies and using performance scorecards.

Appendix B
Tradeoffs and Benefits Chart

	Transactional Cost	Benefits	Cons
Instrument Type: Survey / Questionnaire			
Post - Single Shot	Relatively easy to create and implement	Can provide useful information	Validity concerns, may not capture participant changes accurately
Pre/Post	Higher than post only, requires more analysis	More complete than single-shot post only evaluation	Potential Hawthorne effect
Mid	High - Interupts project process	Low	Interrupts process, potential unintended consequences
Implementation Timeframe of Post or Focus Group			
Upon conclusion	--	Potential high completion rate	Feedback possibility skewed by post process emotions
Within 3 months of conclusion	--	--	--
Within 6 months of conclusion	--	Allows Center more preparation time	Feedback may be less reliable due to participant memory
Survey format			
In person interviews			
By Practitioner	High - requires high number of hours	Allows for in-depth feedback - familiar with project history and participants	Potential conflict of interest - participants may be disaffected
By Neutral	High - requires high number of hours	Allows for in-depth feedback	Unfamiliar with project history and participants
By Project Manager	High - requires high number of hours	Allows for in-depth feedback	--
Survey Monkey	Low	Easy to use and analysis	Limits feedback and potential participation
Mixture of Both	Medium	Gives participants options	Complexity resulting from mixed models
Survey Size			
10 - 20 minutes	--	Higher completion rates	Lower degree of information
20 - 30 minutes	--	--	--
30+ minutes	--	Lower completion rates	Higher degree of information
Who Participates			
Participants	Low	--	--
Dropouts	Potentially High in terms of difficulty gathering feedback	Provides important process information for The Center	--
Practitioner	Low	--	--
Community Members	Medium	Creates full feedback which could be useful for The Center	Highly subjective

Instrument Type: Post Focus Groups	High cost for both The Center and participants in terms of hours and intensity	High degree of information, potential for dynamic feedback	Selection issues
Instrument Type: Neutral Observer	Very high cost to Center	Potential for detailed information	Possible instrumentation issues, single point of view feedback
Instrument Type: Case Studies	Very high cost to Center	Potential for detailed information	Delayed usefulness, no systematic evaluation
Questions			
Hard	Low	Potentially easy for participants, easy to analyze	Lack of detailed information presents reliability and sensitivity issues
Soft	Low - Medium		Validity concerns
Transformative	Low - Medium		Validity concerns
Participant Perceptions	Medium - requires detailed participant reflections		Validity concerns
Quantitative questions	Low		Lack of detailed information
Qualitative / Essay response questions	Medium - requires more of participants	Provides detailed information	Validity concerns, difficult to analyze

BIBLIOGRAPHY AND RESOURCES

Literary Resources

1. ADR Program Evaluation Recommendations. (October 4, 2000) 65 Federal Register 59200, 59208-14. Available at: http://www.epa.gov/adr/evalu.pdf
2. Babcock, Julia. Oregon Solutions Evaluation Report. 2009-2010. Internal Document.
3. Balbach, D. Edith. (1999) Using Case Studies to do Program Evaluation. Stanford Center for Research in Disease Prevention.
4. Bingham, Gail. (1986) *Resolving Environmental Disputes: A Decade of Experience*. Washington DC: The Conservation Foundation, 1986.
5. Buckle, L. G., and Thomas-Buckle, S. R. "Placing Environmental Mediation in Context: Lessons from 'Failed' Mediations." Environmental Impact Assessment Review, 1986, *6* (1), 55–60.
6. Church, Cheyanne, Julie Shouldice. The Evaluation of Conflict Resolution Interventions: Framing the State of Play. INCORE International Conflict Research. October 2002.
7. Church, Cheyanne, Julie Shouldice. The Evaluation of Conflict Resolution Interventions, Part II: Emerging Theory and Practice. INCORE International Conflict Research. March 2003. University of Ulster & The United Nations University. Available at www.incore.ulst.ac.uk
8. Cohn, Jeffrey P. Environmental Conflict Resolution. BioScience 52(5):400-404. 2002
9. Conley, Alexander and Moote, Margaret A.(2003) 'Evaluating Collaborative Natural ResourceManagement', Society & Natural Resources, 16: 5, 371 — 386
10. Cormick, Gerald W. (January 2003) Think Piece V: Assessing Mediation Quality. *Draft version.*
11. Cormick, Gerald W. (1996) Building Consensus for a SustainZble Future: Putting Principles into Practice. National Round Table on the Environment and Economy.
12. Dukes, E. F. (2004), What we know about environmental conflict resolution: An analysis based on research. Conflict Resolution Quarterly, 22: 191–220. doi: 10.1002/crq.98
13. Emerson, Kirk, Patricia Orr, Dale Keyes, and Kathy McKnight. 2009. "Understanding Environmental Conflict Resolution: Performance Outcomes and Contributing Factors." Conflict Resolution Quarterly 27(1). 2009. 27-64.
14. Emerson, K., O'Leary, R. and Bingham, L. B. (2004), Commentary: Comment on Frank Dukes's "what we know about environmental conflict resolution". Conflict Resolution Quarterly, 22: 221–231. doi: 10.1002/crq.99
15. EPA CPRC Evaluation Report. December 2008.
16. Fairman, David. (January 1999) Evaluating Consensus Building Efforts: According To Whom? And Based On What? Consensus.
17. Foley, Tony. Environmental Conflict Resolution: Relational and Environmental Attentiveness as Measures of Success. Conflict Resolution Quarterly. Volume 24, Issue 4. 2004.
18. Fung, Archon. Varieties of Participation in Complex Governance. Public Administration Review. December 2006
19. Harmon, Will. (January 1999) Montana Group Tries Scorecard Approach. Consensus.
20. Hedeen, Timothy, PH.D. (September 2002) Using Participant Feedback to Evaluate and Improve Quality in Mediation. CADRE. Available at http://www.directionservice.org/cadre/
21. Holzinger, K. (2001), Evaluating environmental mediation: Results from a waste management conflict in Germany. Conflict Resolution Quarterly, 18: 397–427. doi: 10.1002/crq.3890180407

22. Innes, Judith E. & David E. Booher, (1999) "Consensus Building and Complex Adaptive Systems: A framework for evaluating collaborative planning" 65:4 Journal of the American Planning Association 412-23.
23. Koontz, T. M. and Thomas, C. W. (2006), What Do We Know and Need to Know about the Environmental Outcomes of Collaborative Management?. Public Administration Review, 66: 111–121. doi: 10.1111/j.1540-6210.2006.00671.x
24. Moore, S. A. "Defining 'Successful' Environmental Dispute Resolution: Case Studies from Public Land Planning in the United States and Australia." *Environmental Impact Assessment Review*, 1996, 16 (3), 151–169.
25. O'Leary, Rosemary, Susan Summers Raines. Lessons Learned from Two Decades of Alternative Dispute Resolution Programs and Processes at the U.S. Environmental Protection Agency. Public Administration Review. Vol. 61, No. 6 (Nov. - Dec., 2001), pp. 682-692
26. Orr, Patricia J., Kirk Emerson, Dale L. Keyes. Environmental Conflict Resolution Practice and Performance: An Evaluation Framework. Conflict Resolution Quarterly 25 (3). Spring 2008.
27. Orr , Patricia. ECR Performance Evidence from the Field - Summary (January 2009) (1) U.S. Institute for Environmental Conflict Resolution
28. Patton, Michael Quinn. *Utilization-Focused Evaluation* 4th Ed. (Los Angeles, SAGE Publications, Inc. 2008)
29. Rolph, Elizabeth S., Erik Moller. Evaluating Agency Alternative Dispute Resolution Programs: A Users' Guide to Data Collection and Use. RAND Corporation. 1995
30. Rossi, Peter H. et al. *Evaluation, A systematic Approach* 7th Ed. (Thousand Oaks, SAGE Publications, Inc. 2004)
31. Rothman, J. (1997), Action-evaluation and conflict resolution: In theory and practice. Conflict Resolution Quarterly, 15: 119–131. doi: 10.1002/crq.3900150206
32. Ruckelshaus, William D. From Conflict to Collaboration: Restoring Trust in Government. May 1997.
33. Susskind, Lawrence & Patrick Field. Dealing with an Angry Public: The Mutual Gains Approach to Resolving Disputes (New York: Free Press, 1996).
34. Thomson, A. M. and Perry, J. L. (2006), Collaboration Processes: Inside the Black Box. Public Administration Review, 66: 20–32. doi: 10.1111/j.1540-6210.2006.00663.x
35. Todd, Susan. (2001) Measuring the effectiveness of environmental dispute settlement efforts. 21 Environmental Impact Assessment Review 97.
36. Scharf, Lee. (October 2004) Evaluating ADR Programs. 65 Federal Register 59200, 59208-14
37. University of Ulster & The United Nations University. Available at www.incore.ulst.ac.uk
38. Warters, Bill. PROGRAM EVALUATION TOOLS for Campus Conflict Resolution & Mediation Programs. Wayne State University, Detroit, MI.

Interviews

Babcock, Julia	Oregon Solutions
Brunner, Jay	Past Project Participant
Cormick, Jerry	Principle, CSE Group
Daniels, Betsy	Senior Facilitator, Triangle Associates
Dukes, Franklin	Director, Institute for Environmental Negotiation, University of Virginia
Ehrmann, John	Founder, The Meridian Institute
Gaffney, Michael J.	Interim Director of the Division of Governmental Studies and Services, WSU
Giles, Sarah	National Policy Consensus Center, Portland State University
Hall, William	EPA, Conflict Prevention and Resolution Center
Harrison, David	Senior Lecturer of Public Affairs, University of Washington
Hedeen, Timothy	Associate Professor of Conflict Management, Kennesaw State University
Hummell, John	Director, Oregon Consensus Program at Portland State University
Imperati, Sam	Institute for Conflict Management
Jones, Robert (Bob)	Director, Florida Conflict Resolution Consortium
Katz, Aaron	Principal Lecturer, Dept Health Services UW
McDaniel, Rob	University Network for Collaborative Governance and Ruckelshaus Center Associate Director
McGinnis, Cat	Oregon Consensus
Moore, Deborah	Past Project Participant and Center Advisory Board Member
Orr, Patricia	Coordinator for Program Evaluation Project, US Institute for Environmental Conflict Resolution
Purdy, Jill	University of Washington, Tacoma
Ross, Bill	Ross and Associates
Ruckelshaus, Bill	Madrona Venture Group and Chair, Ruckelshaus Center Advisory Board
Ryan, Clare	Professor: Environmental policy; Urban ecology and management; Conflict management, University of Washington
Smutko, Steve	Ruckelshaus Institute, University of Wyoming
Wheeler, Bob	President, Triangle Associates

Survey Instruments Reviewed

1. The U.S. Institute of Environmental Conflict Resolution (various surveys)
2. Oregon Consensus and Oregon Solutions
3. Timothy Hedeen, PH.D. (September 2002) Using Participant Feedback to Evaluate and Improve Quality in Mediation. CADRE. Available at http://www.directionservice.org/cadre/
4. Elizabeth S. Rolph, Erik Moller. Evaluating Agency Alternative Dispute Resolution Programs: A Users' Guide to Data Collection and Use. RAND Corporation. 1995
5. Franklin Dukes Instrument Example #1
6. Franklin Dukes Instrument Example #2
7. Franklin Dukes Instrument Example #3
8. Anonymously Supplied Instrument #1

General Resources

- Excerpts from various practitioner interviews available at http://www.civilrightsmediation.org/
- Alternative Dispute Resolution Act of 1998
- Evaluation Summit. The Center for Conflict Resolution at Salisbury University.

Resources for Semi-structured snowball interview process

- Beth L. Leech. Asking Questions: techniques for Semi-structured Interviews. PSOnline, aspanet.org. December 2002.
- Michael Kern. Jan 2011.

Survey Design

Seibert, Therese. Designing Surveys That Count, a Workshop. The Community Research Center at Keene State College, 2002.