

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 aggregated¹ 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 of 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, Hedeem 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 joint 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 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 (Hedeem 2002)

Timothy Hedeem, 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 Hedeem’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 (Hedeem 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.