

## 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 rely 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 Duker 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).