![BetterEvaluation logo]

Issues raised by participants in InterAction Forum
August 2011

As part of the 'Impact Evaluation in Action' session at the 2011 InterAction Forum 2011, participants were invited to contribute towards a discussion about impact evaluation in the nonprofit sector. This workshop was Moderated by Penny Hawkins, Senior Evaluation Officer, The Rockefeller Foundation and session speakers included Carlisle Levine, InterAction Evaluation and Program Effectiveness Working Group Co-Chair, Veronica Olazabal, Executive Secretary for Monitoring & Evaluation, United Methodist Committee on Relief (UMCOR) and Patricia Rogers, Professor of Evaluation, Royal Melbourne Institute of Technology (RMIT).

Participants in the session were asked to consider a recent impact evaluation they had been involved in, and to identify one thing that worked well, and one remaining challenge. The 150 items produced in this 'Sharpies and Sticky Notes' exercise are shown below, sorted into the 7 different components of evaluation that are used to structure the methods in the BetterEvaluation site – Manage & Engage, **Define**, Frame, Describe, Understand Causes, Synthesize, and Report & Support Use.

One of the most striking aspects of the items is the strong emphasis on issues related to evaluation management – in particular, working with partner organizations to leverage the different resources needed for effective impact evaluation, including technical expertise and local credibility to open access to field sites.

The list of items will be helpful in informing the Impact Evaluation Guidance Notes, currently being developed by InterAction, and in further developing material for the BetterEvaluation site. They demonstrate the value of initiatives such as these which seek to share learnings across organizations about effective strategies for evaluation.

MANAGE & ENGAGE

Effective management of an evaluation, ongoing evaluative activity, or the evaluation function within an organization is essential. It includes the following tasks:

- *Ensure appropriate governance of the evaluation*
- *Identify and engage intended users and other important stakeholders*
- *Ensure the appropriateness of those conducting the evaluator*
- *Determine and secure existing and additional resources*
- *Develop Terms of Reference*
- *Ensure an ethical evaluation*
- *Ensure a quality evaluation*

Worked well:

1. Create shared value/goals/expectations in impact evaluation

2. Cross institutional teams: * pull from multiple levels of expertise, * refinement of approach, * broader institutional buy-in because of cross institutional team not HQ driven
3. Enthusiasm, interest and recognition that impact evaluation is important! Small group dedicated to regularly working on impact assessment
4. The 3 key stakeholders – 1) program leadership, 2) evaluation team and 3) funders – spent sufficient time up front to allow full discussion and agreement on evaluation questions, understanding what would be learned and would not, and trade offs/open issues that would need to be revised as the evaluation played out.
5. Convening – groups have been willing to come together and discuss issues and agree that there is a need. Structure for discussions
6. Cross institutional teams: * pull from multiple levels of expertise, * refinement of approach, * broader institutional buy-in because of cross institutional team not HQ driven
7. Developed an online tool tailored to our organizations' way of doing business – contract software/GUI development, – solicit input/feedback from staff for ongoing improvement of the system
8. Enthusiasm, interest and recognition that impact evaluation is important! Small group dedicated to regularly working on impact assessment
9. Finding partner with common vision
10. Good indicators – well thought out, lots of input (Also reported in DESCRIBE)
11. (User of external evaluation) Useful outsider perspective help change some ideas
12. Contract for independent external evaluation with academic professionals
13. Found a data collection expert who could adapt to field realities – he did field testing to verify which methodologies could work
14. My agency has recognized that conducting sound, appropriate IE is a different skill set than managing good humanitarian/development projects so we've hired a team of specialists/academics-social scientists familiar/skilled in IE to design and work with our country teams. They're rolling out training, back stopping etc. to ensure solid IEs.
15. Recruiting a sharp Fullbright scholar to do the field work and working with them to develop the methodological tools required
16. Good involvement of program participants in project and evaluation design, implementation, analysis, reporting and use.
17. Involving a variety of stakeholders in the design, evaluation and dissemination for increased ownership of results
18. Involving multiple stakeholders in evaluation: domestic universities, school and govt actors on the ground. This enables efficient evaluations.
19. Multigroup collaboration on primary data collection tool
20. Participatory/stakeholder involvement in evaluation (all stages)
21. Partnership and relationship building between field and HQ
22. Team were experts in subject matter and country
23. Working in collaboration with other organizations and the community to decide on the real goal of intended impact has made the evaluation process more poignant. Meetings, focus groups, think tanks/brainstorming (Also classified under DEFINE)
24. Extreme attention to logistical planning to ensure enough time (and $) for the necessary process to be done well and communicated to stakeholders
25. Training new in-country research assistants and enumerators (merging their local expertise with data collection techniques)
26. Working in conjunction with partners to reach broader audience and gather more and better data – leveraging scarce resources
27. Collaborating with other organizations who have better knowledge on how to do impact study and evaluations
28. Putting in place the structure to actually start doing M & E – budget and staff
29. Funding for impact evaluation
30. Clear evaluation Terms of Reference with qualified but diverse evaluators carrying out the evaluation
31. Truly randomized assignment of communities between treatment and control arms with a sufficient number in each group largely due to good partnerships with a research University. (Also classified as part of UNDERSTANDING CAUSES)

Challenges:

32. (User of external evaluator) – Evaluator did not come on site, worked remotely (interviews and paper documents)

33. Need adequate budget, staff person hours and time to undertake impact evaluation well.

34. (User of external evaluator) – Evaluator did not come on site, worked remotely (interviews and paper documents)

35. Capacity building and standardization in a global network of autonomous organizations

36. Capacity for data collection to ensure quality and accuracy of information (Also categorized under DESCRIBE)

37. Common understanding between evaluators and policymakers/program managers

38. Cost is the number one challenge

39. Getting to "Yes' – difficult to get agreement around specifics, – i.e. measurements for success. (Also categorized under DESCRIBE)

40. Lack of appreciation on why the data is needed on the part of practitioners on the ground

41. Household surveys to measure impact indicators are resource intensive (time and money)

42. In collaborative projects (multi-agency) getting a common understanding of evaluation

43. Insufficient resources to conduct comprehensive/long term evaluation

44. Resource constraints

45. Resource limitation of implementing partners (particularly staff, time and communication)

46. Resources, especially for post-program/project evaluation

47. Scarce resources (human and financial) to collect data

48. Separating impact evaluation from strategic planning

49. Cost involved to carry out rigorous methods

50. Inadequate capacity to conduct evaluations

51. And resource constraints of course

52. Evaluation capacity of field practitioners (staff, local partners)

53. Finding resources be it cash or personnel to carry out the impact study of various programs

54. Funding/finding partnerships for impact evaluation

55. Managing the dynamics/relationships between NGOs (practitioners) and universities (academics)

56. Measuring impact when a program contract has closed (Also categorized under DESCRIBE)

57. Not enough time/funding to measure impact. (Also categorized under DESCRIBE)

58. To train the staff, the field staff to open their minds to embrace the new documentation required, collect the information in time

DEFINE

It is important to develop an agreed definition of what is to be evaluated. This can require ongoing review and revision it includes the following tasks:

- *Develop a brief description of what is to be evaluated, including its objectives and intended impacts and rationale*
- *Identify potential unintended results (positive and negative)*
- *Develop a theory of change/logic model/program theory of how the program activities are understood to produce or contribute to the intended impacts*

Worked well:

59. Building in impact indicators (not just outcome indicators) into the program design. This would also include unintended, but feared, negative indicators/results/consequences

60. Improve program logics

61. Use of theory of change

62. Good evaluation results – Long-term path of intervention. How commitment/sustainability was built with various actors over time

63. Good evaluation results – captured the relationship building that led to replication and intersectoral multifaceted commitment to project ends. (Also classified as part of UNDERSTANDING CAUSES)

64. Working in collaboration with other organizations and the community to decide on the real goal of intended impact has made the evaluation process more poignant. Meetings, focus groups, think tanks/brainstorming (Also classified under MANAGE)

Challenges:

65. Tried to evaluate impacts of a program without established or clear rules on who receives the benefit/intervention

66. Whether and how to measure the impact of the network or the individual practitioners

FRAME

Before designing the evaluation, it is necessary to decide on its purpose, the key evaluation questions it is intended to answer, and to clarify the values that will be used in the evaluation. Tasks include:

- *Determine purpose*
- *Develop key evaluation questions*
- *Clarify values*

Worked well:

67. Use of reflecting on Peace Practice Methodology to determine strategic and logical project/programme progression – 2 x 2 matrix and criteria for effectiveness

68. Worked well – Establish credibility with the communities we work in

Challenges:

69. Finding a balance between what the results actually turned out to be and what has been promised to the donor, and communicating a truthful and transparent set of outcomes that will satisfy the donor. Issues to consider include: (1) original indicators of success could have been tailored to donor needs, (2) donors are not as interested in "learning" anymore, but are heavily focused on "impact", (3) donors expect to see impact in too-short a period of time – not feasible. (Also classified under REPORT & SUPPORT USE)

DESCRIBE

The design of an evaluation sets out how data will be gathered/retrieved and analysed in order to describe the activities undertaken by an intervention, its outputs, outcomes and impacts, and important contextual factors, including activities by other organizations that contribute to impacts. Tasks include:

- *Sample*
- *Collect/Retrieve data*
- *Use specific measures and indicators*
- *Combine quantitative and qualitative data*
- *Look for patterns*

Worked well:

70. Combining qualitative with quantitative

71. Full participation by the beneficiaries to provide requested information through formal interviews, surveys in written form

72. Good indicators – well thought out, lots of input (Also reported in MANAGE)

73. Use various qualitative, quantitative and tech evaluation to collect data and monitor impact

74. A qualitative study capturing extensive life histories. It was a way of capturing info the field workers knew, confirming programmatic changes on the ground, and giving good tool/data for analysis

75. By working with women's advocacy groups, we have been able to increase participation of women in our programming and measure how many women participate

76. Case study sampling to describe changes to lives of children and families

77. Collecting and graphing epidemiological data about what were the human health outcomes

78. Considering evaluation designs that use existing administrative data

79. Focus groups

80. Getting the patient info as they come to the clinic the first time

81. Good baseline data on disease surveillance infections disease transmission (Evaluation of the impact of the Disease Surveillance Initiative in the lives of people by the Mekong Basin, Asia)

82. Having hub facilitators in each region worked well because they maintained contact with beneficiary communities that had completed projects early in the program and were able to get back in touch with them to arrange focus groups for the final evaluation

83. Involving the field in the formulation of indicators

84. Keeping track of numbers – well organized on a spreadsheet, quarterly/regular reports from field sent to information manager/coordinator who keeps them well ordered.

85. Looking at similar measures used by other organizations to build off their experience

86. Met with and got governmental level permission to access/enter orphanages to do sampling and testing

87. Our country teams work hard to contextualize questions and support valid translations in their local environments

88. Working with trusted local partners to collect data in a local community rather than having a complete outsider try to collect and interpret data

89. Client's Centre 1. Set objectives 2. Achieve objectives – specific indicators to measure it. -> short and long term impact

Challenges:

90. Developing metrics/indicators to measure real gender empowerment and not just capacity strengthening

91. Communicating M & E tools to staff and familiarizing them with the terminology and helping them conceptualize their program with M & E in mind – e.g. identify outcomes, progress markers, indicators etc

92. Designing and maintaining quality, ongoing monitoring systems that feed into the evaluation

93. A measurement statistic for determining a realistic rise in living standards – i.e. does increase in income and number of people working in a region translate to higher living standards?

94. Capacity for data collection to ensure quality and accuracy of information (Also categorized under MANAGE)

95. Consistency in data collection and observation on social impacts

96. Determine sample size.

97. Disconnect between the timeframes for the outcomes of interest and the timeframe for the work being conducted/the program timeframe

98. Distinguishing between "outputs" and "outcomes" (impact)

99. Getting to "Yes' – difficult to get agreement around specifics, – i.e. measurements for success (Also categorized under MANAGE)

100. Lack of consistency in data collection

101. Measuring global advocacy campaigns

102. Measuring capacity building since it's hard to quantify and it's never "done".

103. Collect baseline data that methodologically is geographically ethnically etc representative before beginning our interpretation (in this case on nutritional status of 1,000,000 children in Chinese orphanages)

104. Aligning collaborative partners around common tools and methodology (especially across multiple cultures)

105. Baseline data (challenge that there wasn't any or if it was available, wasn't useful).

106. Changing indicators – early ones selected at the beginning of project don't /are less relevant later on – and new ones emerge as necessary to track

107. Choosing appropriate incentives during surveys to compensate people for their time responding to questions.

108. Creating indicators in areas that are new to use (such as gender)

109. Difficult to measure 'ripple effect'

110. Difficult to measure increased participation of persons with disabilities in our programming.

111. Difficulty with baseline data.

112. Field logistics

113. Follow up training evaluations 30-60-90 days or a year afterwards. Difficulty accessing trainees afterward.

114. Following up with the patient after the initial visit – phone numbers changes, addresses moved, etc

115. Indicators

116. Lack of access to statistics from sponsoring agencies that might serve as indicators of principles being applied.

117. Lack of baseline data (data was flawed).

118. Lack of relevant baseline data

119. Long –term process – difficult to capture in interview methodology because it went beyond the timeframe of the defined project on both sides

120. Poor language translation – did not always ask the exact questions we wanted

121. Quality and validity of data collected

122. Tendency to emphasize outputs (number of people trained) over outcomes (seizure)

123. Developing a simple but effective system to capture outcomes data for a number of years

124. How to decide what key performance evaluation indicators to report on (How to measure impact without standards)

125. Measuring impact when a program contract has closed (Also categorized under MANAGE)

126. Not enough time/funding to measure impact. (Also categorized under MANAGE)

127. On a long program that has gone through several CoPs and program changes, it can be difficult to collect standardized baseline data in all communities.

128. Scope – challenges in data collection

UNDERSTANDING CAUSES

One of the features which distinguishes impact evaluation from other types of evaluation is its focus on understanding the cause-and-effect relationship between the intervention and the outcomes and impacts. Tasks include:

- *Check results match the program theory/logic model*
- *Compare results to counterfactual*
- *Investigate alternative explanations*
- *Investigate complementary explanations*

Worked well:

129. Even with little information there was evidence of impact

130. Good evaluation results – captured the relationship building that led to replication and intersectoral multifaceted commitment to project ends. (Also classified as part of DEFINE)

131. Truly randomized assignment of communities between treatment and control arms with a sufficient number in each group largely due to good partnerships with a research University. (Also classified as part of MANAGE)

132. Use of control group vs. experimental group to determine impact on beneficiaries

133. Working with a comparison group instead of a true control group if it's not possible get a true control group. Adds more rigor than no comparison at all.

Challenges:

134. Adapting message on "rigor" and RCTs to our work – which measures change in whole communities. So we had an entire community to be an intervention or a control and somehow we have to finesse the message to communities that are randomly assigned to be controls that "you just have to wait 3 years" until their turn to get the intervention.

135. Capturing and reporting on qualitative data in a meaningful way that helps to show impacts

136. Designing Impact Evaluations that do not compromise the integrity/purpose of our programming. Balancing the "research" with the delivery of work in the field.

137. Discerning attribution, particularly how to estimate the share of attribution across different partner agencies. How do we give credit for instance to the donor? What share of attribution do they get credit for?

138. Dynamic of success and collective commitment made it difficult for any one organization to clearly claim credit for results against budget investments

139. It was a qualitative evaluation with a relatively small sample size, so attribution was a concern.

140. Presenting rigorous quantitative story about impact

SYNTHESIS

There are three different types of synthesis that might be needed:

- *Synthesis of evidence about a single intervention*
- *Synthesis of evidence of different components of an intervention (e.g. synthesis of projects in a larger program)*
- *Synthesis of evidence from different evaluations*

Worked well:

141. Compared several interventions and their comparative impact to learn which factors have the strongest influence on outcomes. Found that some of our interventions were not building blocks of change or as influential as we thought.

Challenges:

142. Professional/expert judgment outweighs data – bias of SME technical evaluators

143. Translating numbers and data into real impact statements about quality,

REPORT & SUPPORT USE

In addition to managing and framing strategies to support use (such as identifying and engaging intended users and clearly articulating purposes), there are specific tasks involved at the reporting stage of an evaluation and afterwards:

- *Choose relevant report format*
- *Ensure report is accessible*
- *Support use*

Worked well:

144. Communicating success based on success case stories

145. Evidence for policy

146. Identify particular ways to implement the intervention and achieve effect – Bangladesh Stepmothers

147. Using evaluation results with various audiences

 Challenges:

148. Deliver results with high staff turnover rates

149. Culture of evaluation (cut of funds, lack of will to be accountable for one's activities, to show there are mistakes)

150. Finding a balance between what the results actually turned out to be and what has been promised to the donor, and communicating a truthful and transparent set of outcomes that will satisfy the donor. Issues to consider include: (1) original indicators of success could have been tailored to donor needs, (2) donors are not as interested in "learning" anymore, but are heavily focused on "impact", (3) donors expect to see impact in too-short a period of time – not feasible. (Also classified under ENGAGE & FRAME)

151. Getting the "last mile" – full disseminations, verification by participants in the process done well and fully – i.e. according to HAP principles/

152. Knowledge transfer is time consuming

153. Presenting data collected from an impact evaluation in a way that is most useful to implementing partners so that they can use data/analysis not just to look back at past projects but also to plan for future/current projects

Are there other important areas of an impact evaluation that have worked well – or had ongoing challenges?

 - This information was adapted from a blog posted by Prof. Patricia Rogers on the BetterEvalaution Website in September 2011.