

Lessons about multi-agency evaluation

Introduction – this is a work in progress

The Asian Tsunami Evaluation Coalition (TEC) is one of a very few number of **multi-agency of humanitarian action**. This document contains lessons from the TEC to inform similar evaluations in the future. Most of these lessons were identified during a short learning review meeting in February 2006 between some evaluation team leaders and staff, secretariat staff and the management group. Team leaders contributed the majority, but not all views during that meeting. The first version was circulated to the TEC stakeholders and their comments were incorporated into this second version. However, the lessons contained in this document have not been debated. ***Thus it should be read as a compilation of views, with some views enjoying more consensus than others.*** As such, this document should be used to inspire objective discussions about planning, rather than describe what planning must occur. The information has been roughly grouped in a sequence of steps (from preparation to start-up, to data collection, analysis and reporting) to make it easier to apply when having those discussions.

Preparation – defining the scope of the evaluation

1. As a very large evaluation, with a very large topic, the TEC represented a promising intellectual challenge for some. It was exciting to be involved in an event that could represent a 'sea change' in the international humanitarian assistance regime, similar to the impact of the multi-donor evaluation of the response to the Rwanda 1994 crisis. The TEC has the potential to stimulate meaningful debate and it was important for some to be involved in an interagency meta-evaluation with donor involvement. Moreover, the media exposure was exciting for the participants, as usually evaluations do not enjoy such attention. It was also attractive that the TEC was intended to reduce the numbers of field evaluations. The TEC 'venture' was an exciting and even daring one conducted 'against the odds'.
2. The **scale** of the Tsunami crisis challenged the teams. The wide geographic area, the number of donors, number of responding agencies, and the number of affected countries inevitably led to superficiality, particularly because the TORs were quite detailed. That the evaluation was organised thematically (based on weaknesses of the system) was interesting and unusual. This **framework challenged** the teams in two ways: first, the themes were by definition difficult to investigate (because they were systemic weaknesses); and second the themes overlapped. Many seem to feel that the **TORs were overly ambitious**, (leaving the teams with little time to make an in-depth analysis of key issues because of the pressure to comply with the extensive list of issues mentioned in the TOR). It is worth considering an agreed short list of key joint evaluation topics before the next disaster, perhaps **4-5 broadly applicable transversal themes**. In addition to enabling more profound study of a smaller set of issues, this would enable more rapid launching of the evaluations, and provide the possibility of real-time evaluations (although a balance is needed between influencing the response in an RTE and retaining independence). More than one person felt that a quick launch of the evaluation is more important than delaying the evaluation to favour the production of detailed TORs.
3. It was also proposed that **future meta-evaluations might be appropriate to the issues and problems of the particular disaster**, rather than an externally imposed framework. One person elaborated: "...My sense is that there may be inadequate fit between some of the issues/themes the TEC focuses on and those that are of concern in the affected regions regarding post Tsunami relief, recovery, and development. Some would have preferred to see a stronger and comparative (affected) country and conflict focus in the studies which would also have enabled more concrete feed back into on-going reconstruction operations..." Perhaps a participatory process involving field staff, communities and governments to select the topics would produce more practical conclusions that target audiences are more likely to use.
4. Moreover, it is also recommended that the **target audience be clarified and agreed** at the beginning of any future joint evaluation.
5. One desired outcome of the TEC was to reduce the overall number of evaluations. Yet for some agencies, donors are continuing to ask for reports and evaluations on how their funding was spent. Thus, it would appear that while the TEC addressed recurrent systemic weaknesses, it has **not reduced the need for agency evaluations**. Perhaps, until multi-agency evaluations examine results (i.e. did the humanitarian response make a difference to those affected by the disaster), or until more transparency about results exists, operational agencies will be obliged to separately evaluate and account for their actions.

6. In a related point, one person felt that there was **insufficient consideration of impact** in the TEC. "...Process (coordination, needs assessment, funding, etc.) is important, but what matters is outcomes...[and] I have never come across an evaluation where so much has been spent, so much discussed, and so many conclusions drawn with so little reference to, or evidence on, impact and outcomes."

Preparation – structure, governance, participation and staffing

7. **Thematic Steering Committees** were a source of strength, although some of them could be improved next time with a broader representation from policy people, development practitioners, and affected communities. In one case, the configuration of one commissioning agency chairing the process, with technical advisors attending, proved to be optimal. Moreover, the dual and potentially contradictory role of steering committee members (as representatives of agencies that were being evaluated, and as advisors to ensure the impartiality of the evaluation) needs to be recognised and addressed in future evaluations.
8. Some Team Leaders felt the TEC **Core Management Group** could have involved Team Leaders more in its decision-making. For example, the CMG would have made a more informed decision over the dissemination strategy if they had considered ideas generated in the team field trips that had been previously undertaken. Some reflection should occur the next time to ensure that the Management Group membership is drawn from a broad constituency, and to ensure that its size and ways of working help it to make effective decisions.
9. One study noted an underlying paradox that the TEC process has an audience and rationale that is primarily **northern and donor-centric**, yet its goals were to change field level practice. With broader participation on steering committees perhaps this paradox could be partially addressed. Perhaps the question for reflection when the next evaluation is planned is simply **to whom is the evaluation accountable?**
10. Several people were concerned that **TEC meetings have been largely held outside the Asian regions** that were primarily affected by the Tsunami disaster. Some team members who live and work primarily in the affected countries and regions have not been able to participate in any of the TEC meetings most of which were held in Europe where the structure, process and output of the TEC was discussed and evolved. More frequent meetings held in the disaster context would enable affected countries to gain from the evaluation and improve reconstruction.
11. One person felt that in the Tsunami relief operation, accountability to donors has apparently trumped accountability to beneficiaries and the countries affected by the disaster, and that the TEC may also reflect this. In the context **the number of experts from Asia** on the various TEC evaluation teams would also be a matter of concern since many of the affected countries have high levels of expertise. Multi-agency evaluations can be an opportunity to develop capacities in evaluation and development research, but it requires proactive invitations to institutions, and academics from the affected regions to participate.
12. However staffing evaluation teams with **consultants** remains a difficult challenge. Often the selection of consultants for evaluation teams is based on who is available, rather than who might be best for the job. Senior consultants have enormous amounts of experience, but are booked in advance, often in predictable patterns around the calendar year. It is recommended to select consultants early.
13. It is acknowledged that there are strengths and weaknesses with both **centralised and decentralised multi-team joint evaluation structures**. One person felt that without experience in both centralised and decentralised mechanism, it would be difficult to know which is more effective. "... By choosing the latter modus operandi with separate team leaders it became obvious early on that inter-team working was going to be very limited. Whether this has compromised the TEC process or has enhanced it remains to be analysed."

Start up – navigating bureaucratic layers and issues of timing

14. It would appear that most felt the Team Leaders **entered the process too late** (in September). While the Tsunami response rolled on, the TEC was in gestation for 7 months until suddenly the teams were recruited and parachuted into the crisis. In retrospect there was a long delay from February (when the first agreement about the TEC was reached) to June (when momentum and funding was secured). It is recommended that in the future, a principle be established that a lack of funding is not sufficient reason to delay the launch of a joint evaluation.

15. There is much planning that could be done in advance, such as **anticipating funding needs**. For example, measures could be taken to include evaluation funding into flash appeals.
16. **Contracting and tendering procedures** within agencies were challenging, particularly the need to parcel out the studies to avoid running into financial ceilings that prolong the tendering process. This led to fragmentation of the studies and teams. A more smooth and flexible contracting system is needed, however the trends are moving to more complicated systems. More time at the beginning to negotiate the bureaucracy or allocate separate contracts to one team should be planned for.
17. It was a frustrating start for Team Leaders in the TEC. Usually evaluation team leaders are independent, whereas in this instance some were 'merely players' with limited scope for decision-making. Some felt tensions or contradictions between the role of the CMG and their role. The TORs were felt by some not to be sufficiently clear, there were too many agencies to satisfy, and for some **too many "cooks in the kitchen"**. Yet at the same time, while the TEC had frustrating bureaucratic layers to manage the range of 50+ participating agencies, the backing and support from those agencies was valuable. Without agency support, the TEC would not have been as successful.
18. In addition to a late start, several people felt the time allowed for the **preparation phase** was insufficient. There wasn't enough time allowed for document reading and conceptual interaction between the team members for the design of a work plan and strategy. In some instances therefore, more than a desired amount of improvisation was necessary.
19. The **media pressure** of the 1-year anniversary created an unrealistic timeframe for the TEC, and it would appear that most felt an undue degree of pressure as a result. Perhaps this was the reason that insufficient time was given to the teams to prepare (i.e. conduct background reading etc.) before their deployment to the field. One person felt a major issue throughout the fieldwork stage was the constant pressure of producing 'messages' for the one-year anniversary - which in the end were not published.
20. "... There will always be a (valid) pressure to complete humanitarian evaluations quickly before information is lost. However, issues being inadequately thought through before the evaluations were 'launched' could be offset in part by including highly experienced field evaluators in the early planning stages, to discuss issues such as with setting dates and time deadlines; the practicalities of so many teams working together as part of the TEC; harmonization between the different evaluations etc, how realistic reporting deadlines were. **Involving experienced consultants in the early planning** might have kept expectations more realistic..."

Start up - budgets

21. One person felt that budgeting was troublesome, "...Many teams seemed to have almost no limits on funding while others were very short..." Thus issues of participation may actually be simple budgeting issues, with the recommendation that budgets anticipate the need for team members to travel and meet over the whole process of planning, analysis and reporting.

Start up - teamwork

22. With split funding for the different teams, the TEC was fragmented, and it seems that few felt that they were part of a larger 'TEC team'. One person felt that at the start, there was a degree of resistance to cohesion. Interaction between the teams seems to have occurred only between team leaders. One person felt that more **insightful conclusions might result from greater interactions between extended teams** (not just the leaders).
23. One person felt that there was no uniformity across the different evaluations vis-à-vis the Team Leader's role in the selection of the rest of the Team. Some agencies recruited an entire team to work on their theme, while some recruited individuals who were then expected to form a team. Evidently the former creates an easier working environment. **Regardless, more time should be allowed for team building at the beginning of the evaluation.** It seemed that in this instance many of the individuals recruited into the teams knew each other from other shared experiences, which helped overcome some of the structural constraints to teamwork within the TEC structure.
24. Another person felt that **more structured encounters between evaluators and the Steering Committee** should be organized in the very early stages of the evaluation. "...No matter how strong the leader is, teams need to build common understanding from the start and work as a whole. It would be useful in future joint evaluations to strengthen the Terms of Reference of the sector teams."

25. Some people did not favourably view the system of **research associates**. Conversely, one person felt they were vital to the joint evaluation process: "... although their utility is conditional on at least three elements: their work starts at least one month prior to field work (to provide baseline understanding prior to the field reality check); they serve as active and full-time members of the evaluation team, available to delve into research subjects that arise during field work; and their research contributes directly to a final evaluation report and is not conceived as a stand alone piece from the start..." Another person suggested that it might have been smoother if the Team Leader and researchers started together well in advance of field missions. Another person suggested that researchers have longer terms such that they could conduct a desk study on the topic prior to the rest of the team coming together, and then be constantly available for on-call requests from the field as was needed.

Data collection – methodology and planning

26. Creating a 'diary' of events from the first day of a new disaster creates a **history of the humanitarian response** operation that helps evaluators when they eventually do begin their work
27. One person felt that "...the **evaluation methodology** should be developed at least in a provisional format before venturing out to the field. There should be some agreement on the methodology among the team members and other key stakeholders who commission the evaluation..."
28. It is worth considering **whether primary data collection would be necessary** in future joint evaluations. Instead of running in parallel to agency evaluations, perhaps the next joint evaluation could be a true meta-evaluation and wait for the inevitable large number of evaluations to occur. One person suggested that primary data collection effort could be improved by conceiving it as a parallel effort to the evaluation teams, conducted by an independent data collection team. The results would then be available to all teams simultaneously for use in their thematic analyses.
29. When there are several evaluation teams, some unavoidable **overlap in the terms of reference** for the teams will occur. One person felt however that it would be important to avoid 'unproductive duplication' within the work of the teams. A second consequence of overlap is 'responder fatigue' when several TEC teams visit the same respondents. Often the responders do not see or appreciate the somewhat subtle differences in emphasis and focus that the different teams have and often view it as the same story being repeated.
30. One person suggested that if future evaluations cover more than one country, they might **not need to occur at the same time in all the countries**, thus reducing pressure on the TEC management. Timing should also consider long cultural/religious celebrations that could have an impact in the operations and results of the evaluation work (for example, the Month of Ramadan and Idul Fitri celebrations, as it was the case for Indonesia).
31. **September/October for data collection was the wrong time**. In retrospect, the TEC evaluated the relief issues (such as needs assessment and co-ordination) too late; and evaluated the rehabilitation issues (such as capacity building) too early.
32. One person felt that an indication of the rushed timing was that inception reports were written at the same time as the start of the fieldwork. Another person noted pressures to start producing 'messages' before the team had been able to digest the information from its respective country studies and sit together to collectively draw out the picture emerging.

Data collection – local counterparts and stakeholders

33. **Local resource/liason people** are essential. They improve the efficiency of the process and expedite the processing of interview information and additional reports that were not available from home base. One person suggested that at a deeper level, it is presumptuous to assume that fieldwork in a given nation can be insightful without, at the very least, the participation of 2-3 nationals. "...Their work to bring the team crucial understanding of cultural contexts transcends simple translation and logistics and cannot be underestimated..." Recruiting national counterparts needs to be started early and this element may be one that the Secretariat could organize, supplying to evaluation teams updated lists of available qualified candidates who are standing by and interested. Two weeks before fieldwork begins is too late to recruit national counterparts for evaluation teams. One person felt the TEC had tremendously able people working on the project because the local country team leaders were well-connected and good networkers.

34. Accordingly, sufficient time should be allowed to agree a division of labour between local and international evaluators; for training of local evaluators to take place; and for the gathering of information by local evaluators prior to the field visit of international evaluation experts.
35. Very important for the smooth running of field visits is **buy in from local offices**. One person felt this was not evident in a couple of places. Despite an initial visit, local offices were not 'on board' which had the effect of further compounding the shortage of time as teams had to spend time drawing stakeholders into the process. Involving country offices in the very early planning stages, again, would go some way to improving local buy-in. Also a clear mandate for what could be expected by all stakeholders from the initial visits would have been helpful, and minimized their unmet expectations.
36. One person recommended that a much more clear sighted planning process **drawing on experienced consultants** in the field would mean more realistic expectations of the teams and a more realistic view of the implications of fielding so many consultants. From the point of view of local stakeholder attitudes a more comprehensive process of drawing people in from early on and soliciting their opinions could have smoothed things for teams arriving in the field.

Data collection – support from the Secretariat and host agencies

37. The **resource CD** produced by the TEC Co-ordinator was viewed as a fundamental resource, (proved especially valuable where internet connections may have been limited in the field), but one person felt there were too many versions. The document "summary of other evaluations" done by the **Secretariat** was well done and appreciated. Knowledge management and co-ordination from the Secretariat was successful and the back up and support from John and Rachel was immediate, professional and appreciated. They had clear roles, and were well organised. The website also was appreciated.
38. This was a geographically disbursed evaluation in multiple countries with quite different contexts. That each evaluation team was managed by a **host agency** was invaluable in organising all the travel and logistics. Yet, being too closely identified with the hosting agency challenged some teams to maintain perceptions of independence.
39. One person commented that to absorb even part of the 8000 plus documents available as a result the Indonesia field visit was complicated without a proper desk review and adequate logistics in place (problems with cars, phones and translators).

Data collection – coordination between TEC entities

40. One person felt **better co-ordination between field teams** would have helped to share information on what was each team doing and to avoid overlapping. For example, the Local Response Team was not informed/invited to the introductory meetings of the TEC in Jakarta and Banda Aceh. They got to know that other TEC Teams were in Banda Aceh because, by chance, they and another TEC Team used the same local partner for field data collection. "...When I finally could contact a person from another TEC Team and introduced myself, she told me she did not know there was a Local Response Team!" In addition to avoiding duplication, better co-ordination would enable more substantive preparatory discussions between teams with overlapping subject areas.
41. When deployed there seemed to be **too many TEC field people**. Although the TEC was supposed to reduce field missions, it didn't appear to be successful in this regard. Three teams were deployed in September, yet it was impractical for the teams to work together, giving an impression in some field staff of excessive numbers of evaluation teams. It was perhaps too idealistic for the teams to be in the field at the same time.
42. In Indonesia, **TEC field workshops** were felt to be unnecessary by some Team Leaders, because the teams could have run these workshops themselves. Government counterparts and interviewees, while interested in participating in the evaluation, were confused between the evaluation teams and the TEC Secretariat. It probably was unnecessary to try to explain the structure of the TEC to interlocutors.
43. One person wrote that DANIDA served as a real example of a donor working to get things done. "Niels is almost unique and deserves a heartfelt round of applause."

Data collection – obtaining sensitive information

44. One person felt that seeking financial data and expenditures was sensitive and difficult, and that it requires preparation to ensure that such information is requested as part of a program discussion, not an audit discussion.

The suggestion is that perhaps, if government were more closely involved, it would be easier to obtain and the data of higher quality. Another person felt that better assessment of issues like remittances, contributions from communities and local organisations of the civil society need to be included more strongly in future evaluations.

45. The Tsunami response experienced a great number of **audits** as well as evaluations. Field staff seemed to view audits as more important than evaluations, as a bad audit can limit one's career whereas a bad evaluation has less severe ramifications. Was an opportunity for the TEC to collaborate with auditors missed?

Analysis and reporting - methodology

46. Some felt the analysis and reporting should reflect the reality of what was discovered in the data collection phase, rather than what their TORs had outlined. The **feedback from the Secretariat** to the teams on how their reports did not match their TORs was consequently received with some difficulty. Future evaluations could make explicit the flexibility allowed to teams in the development of their findings, and in the role of the Secretariat with regard to the evaluation reports.
47. Some felt time was wasted agonizing over the **structure of the reports**
48. One person observed that the overall slippage of the TEC deadlines from mid-December to end February meant that team members had already started taking up other prior commitments, before country reports were finalized and the regional report written. Inability of the team to meet to discuss findings and propose recommendations was also viewed as a constraint by this person.

Analysis and reporting – outputs and target audience

49. Although the **key findings document** was seen as useful, some felt it required too much effort.
50. One person felt that there appears to be inadequate attention paid by the TEC to dissemination and validation of reports to key stakeholders in the affected countries, reinforcing the perception that the TEC process is too inward-looking and lacks accountability to claimholders in tsunami-ravaged regions. This person feels that it is not too late to remedy this problem, and this concern must be substantively addressed if the TEC is to receive credibility and to embody international best practice.
51. The HRR was not co-ordinated with the TEC. Perhaps the TEC could have produced a more **practical output**, which could have fed into some sort of systemic process like the HRR
52. One person suggested that an overlooked element in the TEC process was the application of the results, and that an additional budget line for lobbying be created. "...Rather than budget only for press releases and communication strategies, we need to write into the concept action plans for how to apply the recommendations. What meetings with which major decision makers and which teams are vital to eventually change X or to propose Y? How to assure that recommendation X will fall on receptive ears and on an agent with not only good will but also capacity to enact change? What must be done to make sure the international community really does better next time?"

Overall observations and successes

53. One person wrote: "...maybe we should instead look for how and why we still managed to muddle through and produce reports that seem to have quite a number of interesting and valid conclusions. To me, it seems that dedication and skill among the various teams is one major factor. The time limit we deliberately imposed on ourselves may have been against better judgement, but probably helped to make the TEC relevant..."
54. It would appear that the TEC had problems, but as one of a small set of pioneering meta-evaluations it should not be expected to solve all problems. Harmonization is a process, of which the TEC was a first step. It is emphasised that joint evaluations will occur in the future, and they need to be refined.
55. **If joint evaluations start with coherent terms of reference that are drafted in a collaborative process, and if they start with clear policies and parameters, they will be more joined up at the end. Perhaps the first step is to decide whether the next evaluation will address the system, or the impact of the response. END**

*Prepared by Sean Lowrie, independent consultant
7 March 2006*