# Issues arising in the development of UNICEF guidance on the evaluation of psychosocial programmes in emergencies

## Wendy Ager

*This paper describes the development of a guide on evaluation, commissioned by UNICEF for their field officers in 2006. The consultation process in developing the content of the guide is discussed, revealing varying perceptions of what is known and practiced in the field in relation to evaluation. Broader findings about evaluation design and methods are also discussed, based on a review by the Mailman School of Public Health. The paper then focuses on specific aspects of evaluation of psychosocial programmes that seem to generate particular difficulties in practice. It considers firstly the conceptualisation of psychosocial wellbeing in the definition of indicators of output, outcome and impact levels. Secondly, case studies are examined to illustrate the challenges of using methods, both qualitative and quantitative, in the context of complex emergencies. Thirdly, the paper considers the tension between 'evaluation as research' and 'evaluation as reality' and discusses the relative values that are reflected in the field.*

**Keywords:** evaluation, psychosocial programmes, emergencies

## Introduction

What makes the evaluation of psychosocial programmes different from evaluations of other programmes? In one sense there is no difference. For an organisation like UNICEF (United Nations Childrens' Fund), standard evaluation criteria apply across all projects. An evaluation of a psychosocial programme, just as any other, would include consideration of the relevance, efficiency, effectiveness, impact and sustainability of the work being undertaken (UNICEF, 2005a).

For the psychosocial field, however, evaluation may also be able to contribute something more. Psychosocial support has become increasingly common, but differing approaches and diverse understandings have led to inconsistencies and concern about conceptualization and methodology (Programme on Forced Migration and Health, 2006). In this context, well documented rigorous evaluations may help to build a much needed stronger knowledge base for good psychosocial practice. This places the findings of an evaluation in a broader context than its immediate benefit to the specific programme it addresses (Patton, 2001). In addition to establishing whether a project is achieving its goals and, in some cases, providing information to develop programming in subsequent phases, evaluation may feed into wider learning, offering findings that can inform work in other situations and settings. This paper discusses the development of UNICEF guidance (UNICEF, 2007) on the evaluation of psychosocial programmes, a task that was commissioned in December 2006.[1] The terms of reference provided by

UNICEF focused on the development of a field friendly guide to psychosocial evaluation in complex emergencies, addressing non-clinical psychosocial programmes (the three layers of the intervention pyramid, not including '*specialised services*', in the IASC Mental Health and Psychosocial Support Guidelines (Inter-Agency Standing Committee (IASC), 2007). It was written for UNICEF, their partners and other organisations working in the field of psychosocial support in thinking through key issues in planning and implementing evaluations. The aim was that it would be a supplement to existing UNICEF guidelines, providing guidance relevant to the specific challenges of evaluating psychosocial programming.

The guide in its current draft has been used in a number of programmes to inform UNICEF evaluation processes and will be further developed and reviewed in a planned period of in field testing.

## Developing the content

The starting point for the guidance was developing a preliminary paper headlining the main topics for the manual. A results based evaluation approach was proposed to reinforce what is currently in use by UNICEF for programme planning and management (UNICEF, 2005b). Consultants of the Mailman School of Public Health, Columbia University, prepared an initial outline of the guidance and an expert workshop was organised to discuss and develop the material. Comments were also invited by email.

This process of developing the content of the guidance involved consultation with UNICEF staff (including specialists in child protection, and monitoring and evaluation), international non governmental organisation (NGO) specialists (both headquarters

and field based staff, drawn from 11 different organizations), academics (with a background in consulting with NGOs) and independent consultants. Commentary ranged over many different aspects of the material, addressing the coverage, level and accessibility of the guidance. The text of the guidance was amended on the basis of these comments. Four issues in particular are discussed here; firstly concerning the '*level*' of the guidance given; secondly relating to the definition of '*psychosocial wellbeing*'; thirdly about the meaning of '*impact*'; and fourthly in relation to the '*ethics*' of comparison groups.

*The 'level' of the guidance.* There were varying perceptions about what purpose the guidance would play for UNICEF field officers and their equivalents in NGOs. If the manual was to be a practical source of help in the field, what might it most usefully include? Should guidance provide the '*how*' or the '*why*' of evaluation? For some commentators, the first draft of the guidance assumed too much knowledge, whereas others felt it was on track. This reflected different views about the capacity of teams that were likely to be available to implement evaluations. Significant training needs in this area were recognised, beyond what could be met by one evaluation guide. Nonetheless, the value of a document clarifying key issues was supported and most commentators advised the importance of providing concrete examples of how to measure changes, specifically in psychosocial wellbeing. The most useful contribution the guidance could make would be to provide examples of evaluation that illustrate good practice.

*The definition of 'psychosocial wellbeing'.* The conceptualisation of '*psychosocial wellbeing*' also received attention. Several commentators were concerned about characterising the field in terms of '*lacking consensus.*' They suggested that this negative reading be

| **Table 1. Three core domains of psychosocial wellbeing for UNICEF evaluations** | |
| --- | --- |
| Skills and knowledge | life skills, vocational skills, conflict management, using culturally appropriate coping mechanisms |
| Emotional wellbeing | feeling safe, trust in others, self-worth, hope for the future |
| Social wellbeing | attachment to caregivers, relationships with peers, sense of belonging to a community, access to socially appropriate roles |

reframed and were looking for a standard definition to encourage practitioners in their work. What evolved in the text was a brief overview of concepts in current usage followed by a more specific proposal that UNICEF evaluations use three '*core domains*': skills and knowledge; emotional wellbeing; and social wellbeing, to determine how well UNICEF's work affects the lives and experiences of children, their families and the communities in which they live. These domains were agreed to cover the central concerns of most psychosocial programming (Table 1).

It is proposed that all UNICEF psychosocial programmes begin to use these three domains to describe their work. In practice, there would be scope to '*contextualise*' these domains within particular settings, but there would also be a consistency and coherence brought to all evaluations. Over time it will then be possible to build a better knowledge base about the field.

*The meaning of 'impact'.* Another debate focused on the assumptions related to the meaning of '*impact*.' The guidance notes that '*project impact is a change in status or behaviour related to stated project objectives*' and anticipates the need to identify such changes for children, their families and/or their communities as a result of programming. The emphasis of much of the guidance, given the need to establish the value of psychosocial programming in supporting post emergency recovery,

is on; '*the long term effects on identifiable populations or groups produced by a project, directly or indirectly, intended or unintended*' (Bamberger, Rugh & Mabry, 2006).

However, other definitions and assumptions abound in the field. It was evident in examples of evaluation reports (Programme on Forced Migration and Health, 2006) that impact could mean, amongst other things; '*the scope or scale*', '*the depth*,' and/or '*the influence*' of a project. Bamberger et al. (2006), examining the use of the term '*impact*,' identified seven different nuances in relation to real world evaluation. We wished to sharpen awareness about these distinctions and the consequences of basing evaluations on such varied assumptions. However, this has continued to be debated through the process of developing the guidance and appears to reflect a wide variety of differing uses in the field.

*The ethics of establishing comparison groups.* There was some concern about the '*ethics*' of establishing comparison groups. This was related to a fear that this constituted a kind of '*systematic deprivation*', because there would be groups that did not receive services at some stage. However, this was challenged by the view that there are also risks of creating harm by delivering an '*untested*' intervention. Providing a comparison group is established sensitively, for example, in the course of rolling out a programme, and then harm should be minimised. For example, it might

be possible to set up comparison groups between children currently enrolled in a programme with those waiting to be enrolled. After all, field interventions are rarely capable of being delivered simultaneously in different areas.

## Common difficulties in implementing evaluations

Certain aspects of evaluation of psychosocial programmes seem to generate difficulties in practice. This section focuses on three issues in particular: the conceptualisation of psychosocial wellbeing; practical challenges of evaluation in the field; and the tension between *'evaluation as research'* and *'evaluation as reality'.*

*The conceptualisation of psychosocial wellbeing in the definition of indicators at output, outcome and impact levels.* A number of challenges appear common to the process of evaluation in the field. The first stems from a lack of clear objectives developed at the project design stage (Programme on Forced Migration and Health, 2006). A fundamental question of evaluation such as *'has the project achieved its objectives?'* becomes a major hurdle without a clear articulation of those objectives. As a consequence, projects may be merely characterised as *'helpful'*, lacking ways of determining their achievements more accurately (Duncan & Arnston, 2004). To counter this, the UNICEF guidelines suggest that an evaluation should always be integrated into the project design. By asking at the planning stage such questions as; *'what impact on children's lives is hoped in the longer term in what we are planning? What would success look like? How would we measure this?',* programme objectives are then clarified and in the process, evaluation processes begin to take shape.

Secondly, as indicated earlier, there are different usages of the term *'impact'*, and in addition to this, the terms *'outputs', 'outcomes'* and *'impacts'* may often be interchanged. When this happens in evaluation reporting, it can be difficult to understand what the evaluation has found. For example, evaluations sometimes feature outputs such as the number of children involved in activities and what those activities were and incorrectly describe them as impacts. The working definitions used in the course of developing the UNICEF guidance were as follows:

1. Project outputs are the immediate accomplishments of the project.
2. Project outcomes are the measurable or observable results from a project, based on the stated project objectives.
3. Project impact is a change in status or behaviour related to stated project objectives.

However such definitions remain under scrutiny, given the widely differing use of such terms noted earlier. It is hoped that clarity and consensus on meaning will emerge from processes of field-testing and sharing the guidance with other organisations.

A third challenge relates to the specific difficulties in defining measures for changes in psychosocial wellbeing at these three levels. It seems that indicators at the output level are more straightforward to identify than at the outcome and impact levels (Adjukovic, 2008). In the UNICEF guidance, the following example was used to demonstrate how indicators could be used. The project described is using expressive arts to address the psychosocial needs of youth who have returned to their home districts after a period of displacement as a result of armed conflict. Table 2 shows one example from each level.

Moving to look more specifically at how to prompt indicators that cover the breadth of psychosocial wellbeing, the guidance

**Table 2**

*Output objectives:*
Community based arts groups established in all districts and delivering planned curriculum to youth 12–16 years

*Indicators:*
No. of groups established
No. of expressive arts sessions delivered
No. of youth reached
Quality activities delivered

*Sources of information:*
Programme log book at each site
Monitoring visits and observation
Project reports

*Outcome:*
Children more socially integrated and better able to express and deal with emotions

*Indicators:*
Increased no. of positive social interactions within specified time
Increase in positive relationships between children and adults
Degree of social functioning according to families

*Sources of information:*
Observations
Focus groups with families
Measurement scale of current relationships
Self report by children

*Impact:*
Children and youth play greater role in community activities and decision-making

*Indicators:*
Increased no. of youth-initiated community activities
Children and youth represented in activities at community level

*Sources of information:*
Interviews with children and youth
Community visits
Records of meetings

suggests that three core domains indicated earlier are used. Where broad impacts on psychosocial wellbeing are targeted, evaluations include indicators in relation to each of the *'domains'* of skills and knowledge, emotional wellbeing and social wellbeing. These indicators may relate to individual or collective aspects of wellbeing, operating across individuals, families and communities. In practice, the activities of a programme may well influence more than one domain. An output, such as a drama group, may result in learning of new skills, increase in children's social networks and enhanced emotional wellbeing.

*Practical challenges of evaluation in the field.* An evaluation of a programme supporting the reintegration of children formally associated with fighting forces in Mozambique (Boothby, Crawford & Halprin, 2006) is used as a case study in the UNICEF guidance to illustrate the challenges of evaluation in the field. In this paper, we look only at assessing the longer term impact, 16 years after the programme ended.

Shown below are the longer term impact indicators, which were used for a well resourced impact evaluation, defined in terms of the three core domains proposed in the UNICEF guidance.

Domain: skills and knowledge
Indicators: ability to be economic providers for household; farming activity; off-farm income

Domain: emotional wellbeing
Indicators: reported symptoms of post traumatic stress disorder, self esteem, sense of community acceptance

Domain: social wellbeing
Indicators: ratings of social functioning by spouses, parents and neighbours

In terms of methods there are two observations from this work that reflect general principles in the guidance. Firstly, it proved very difficult to identify a comparison group of adults for the purpose of the evaluation. There were plans initially to identify a comparison group of adults who had not been beneficiaries of the psychosocial support programme. Given the sensitivity of identifying those who had fought in the war as children and had not received structured support subsequently, it was decided instead to use local norms as a basis for comparison. This was not ideal, but evaluation could conclude that *'those who participated in the programme are doing better than is typical in this area,'* even if it could not conclude *'those who participated in the programme are doing better than those who did not participate.'*

Secondly, *'free listing'* proved to be a particularly useful tool for generating quantifiable information that clearly reflected local understandings of wellbeing. (methodology is described in Armstrong, Boyden, Galappatti & Hart, 2004).

Free listing revealed the following findings that otherwise might have been overlooked:

- Local perspectives on trauma symptoms were similar to those described in the literature
- The cause of these problems was understood differently. Local understandings perceived them as resulting from spiritual contamination that required cleansing.
- Traditional cleansing ceremonies helped both individuals and communities. They engaged spiritual guardians in helping to reintegrate former child soldiers into families and villages
- The most cited characteristic of a *'good and successful adult'* in Mozambique was someone who helped neighbours in need. The next three most commonly cited characteristics were; the ability to

economically support a family household, to be a good spouse and to be a good parent.

*Tension between 'evaluation as research' and 'evaluation as reality'.* The tension that I have characterised as *'evaluation as research'* and *'evaluation as reality'* was reflected in the NATO Advanced Research Workshop (2007) where participants debated the value of evaluation to beneficiaries, practitioners, funders and academics. *Evaluation as research*, on the one hand, values rigorous design and make RCTs (randomised controlled trials) *'the gold standard'*. Much of the research literature in fact indicates a *'hierarchy of evidence'* with systematic reviews and RCTs being deemed the most objective approaches.

However, Glasby & Beresford (2006) argue for a different understanding of what constitutes valid knowledge that should include *'practice wisdom and tacit knowledge of practitioners and the lived experience or human testimony of service users and their families'*. In this vein, therefore, *evaluation as reality* urges the scope of evaluation to address the real questions about the nature of suffering and asks; *'what are we really doing?'* and *'what is the deeper level?'*

What options might there be for evaluation that is concerned with this broader *'evaluation as reality'*? Patton (2001) suggests the formulation of *'high quality lessons learned'*. On the face of it, this notion of *'lessons learned'* seems vague, but in fact it defines an approach that is focused and specific; *'represent(ing) principles extrapolated from multiple sources and independently triangulated to increase transferability'*.

Eight supporting sources are indicated below by Patton:

- Evaluation findings and patterns across programmes
- Basic and applied research
- Practice wisdom and experience of practitioners
- Experiences reported by programme participants/beneficiaries
- Expert opinion
- Cross-disciplinary connections and patterns
- Assessment of the importance of the lesson learned
- Strength of the connection to outcomes attainment

The idea is that greater confidence can be attached to a lesson learned that draws on a wide range of triangulated sources, rather than a narrower range of *'research respectable'* findings. If we accept that evaluation has the potential of contributing to the knowledge base, and lessons are drawn from a wide range of sources as illustrated, then we might be setting up conditions in which practitioners, in particular, feel they have something important to gain from evaluation that addresses *'real issues of suffering.'*

## Conclusion

The process of developing this guidance has mirrored the issues that organisations grapple with in the course of *'doing'* evaluation, such as establishing consistent terminology across a range of players in the field, and balancing the need for valid information with the constraints of time and resources in emergency settings. There are many challenges to effective meaningful evaluation, but many benefits too – not the least for those communities who we seek to support.

## References

Adjukovic, D. (2008). *Development of evaluation indicators in psychosocial projects: Balance between creative challenge and research rigor. Intervention 6(1),* 29-38.

Armstrong, M., Boyden, J., Galappatti, A. & Hart, J. (2004). *Piloting Methods for the Evaluation of Psychosocial Programme Impact in Sri Lanka.* Final Report for USAID. Refugee Studies Centre, University of Oxford.

Bamberger, M., Rugh, J. & Mabry, L. (2006). *Real World Evaluation.* California: Sage.

Boothby, N., Crawford, J. & Halprin, J. (2006). The life outcomes of former Mozambican child soldiers. *Global Public Health, 1*, 87-104.

Duncan, J. & Arnston, L. (2004). *Children in Crisis: Good practices in Evaluating Psychosocial Programmeming.* Save the Children Federation.

Glasby, J. & Beresford, P. (2006). Who knows best? Evidence-based practice and the service user contribution. *Critical Social Policy, 26*(1), 268-284.

IASC (Inter-Agency Standing Committee) Guidelines on Mental Health and Psychosocial Support in Emergency Settings, 2007.

NATO Advanced Research Workshop (2007) Lipica, Slovenia.

Patton, M. Q. (2001). Evaluation, knowledge management, best practices and high quality lessons learned. *American Journal of Evaluation, 22*, 329-336.

Programme on Forced Migration and Health, Mailman School of Public Health, Columbia University (2006). Assessment and Evaluation of Psychosocial Programming for Crisis-affected Children. Funded by UNICEF.

UNICEF (2005a). Monitoring and Evaluation Quick Reference: Extracts from the Programme Policy and Procedure Manual (revised) May 2005. Available at: www.unicef.org/.

UNICEF (2005b). Understanding Results-based Planning and Management: Tools to reinforce Programming Practice. UNICEF.

UNICEF (2007). A guide to the evaluation of psychosocial programming in emergencies.

---

[1] The guidance on the evaluation of psychosocial programmes described in this paper was commissioned by UNICEF. This paper does not represent a formal statement of the views of UNICEF but represents those of the author.

*Wendy Ager, DipSW, MSc, has worked in the NGO sector and with the Universities of Malawi, Dundee and Stirling. She currently works as an independent social worker researching issues of participation and user involvement, and developing guidelines for humanitarian practice. email: Wendy. Ager@gmail.com*