

An introduction to the principles for assessing the quality of evidence

What are the principles and the checklist?

The Principles and associated Checklist is a tool for assessing the quality of evidence collected and used by NGOs to measure, learn from and demonstrate their impact. They can be used to review and quality assure existing evidence (e.g. an evaluation report), and as a reference point when thinking about how evidence will be generated (e.g. to set an evaluation terms of reference). The checklist has been developed by NGOs for NGOs and sets a common standard for rigorous evidence. It builds on NGO values, and sets standards of rigour, which are relevant and practical to the scale of NGO operations and the type of interventions NGOs deliver.

The five principles are: voice and inclusion, appropriateness, triangulation, contribution and transparency. For each principle the checklist has four questions that can be used to test the quality of a piece of evidence. Each question is scored on a 1-4 scale, allowing the user to locate a piece of evidence on a scale of: 1) weak evidence, 2) minimum standard, 3) good practice, 4) gold standard. An overall score and colour (red, amber, light green or dark green) is then assigned to each principle to provide a holistic picture of the robustness of a piece of evidence.

Why has the checklist been developed?

All organisations strive to base their decisions on evidence of what works. NGOs are no different. NGOs want to ensure they are putting their resources into the interventions that can make the biggest difference to the lives of poor and marginalised people. The challenge is that NGOs do not always have a robust evidence base for what works. This is especially the case for areas such as empowerment or accountability, which are difficult to measure. As a result NGOs are not always sure if they are basing decisions on solid evidence of what works, or anecdotal evidence.

Recognising this challenge, NGOs are looking to move forward in a way that is relevant and practical for them. NGOs need a standard of rigour which is appropriate for the methods they use to generate evidence.

Statistical methods have common standards of quality, however similar benchmarks do not exist for the types of qualitative methods, or indeed other, non-statistical quantitative data collection methods that are used by many NGOs in generating evidence. The checklist is an effort by UK NGOs to define a common standard of robustness, taking into consideration the level of resources at the disposal of most NGOs and the types of interventions that they implement.

“This piece feels like the first that really cracks into the challenge of a shared understanding/standard of development effectiveness... I see it as critical to ensuring credibility of our evidence as well as improving our own standards/practice.”

Feedback from a consultation workshop

How has the checklist been developed?

Initial discussions on the checklist were held with a small group of Bond members in Autumn 2011, followed by a consultation with a larger group of 30 NGOs, Comic Relief, DEC and DFID in April 2012. Based on this consultation an initial list of principles and criteria was drawn up and piloted by seven NGOs (CAFOD, Christian Aid, EveryChild, the International HIV/AIDS Alliance, Self Help Africa, Traidcraft and World Vision) between May and June 2012. After a second consultation workshop with 28 NGOs the final version of the criteria was created, containing five principles and twenty criteria. Following discussion with the pilot group and DFID in

July 2012 the decision was taken to develop scales for each of the questions. The scales were drafted by a consultant and agreed at a workshop with the pilot organisations in August.

Moving forward, while the scoring scales will be tested by a number of the piloting NGOs during September and minor adaptations made based on this, on the whole, the Principles and Checklist are complete and will be launched in October alongside the Improve it Framework. Over the coming year we will be working closely with a core group of NGOs to integrate the Checklist into how they design assessment processes and quality assure evidence and using their experiences and feedback to further refine and improve the tool.

How is the checklist structured?

The following section provides a brief explanation of each of the five principles on which the checklist is structured and why it is important to quality evidence.

- 1) Voice and Inclusion:** the perspectives of people living in poverty, including the most marginalised, are included in the evidence, and a clear picture is provided of who is affected and how

Ensuring that the voices of poor people, particularly the most excluded and marginalised, are captured and form the basis of any analysis of what has changed, for who, and how is key to generating compelling and robust evidence. People living in poverty are the best judges of what has changed in their lives and their wider communities and need to be a key source for understanding the contributions of an intervention. The more they are involved in informing the approach to the assessment, analysing and giving meaning to the data and validating findings, the richer and more compelling the evidence of what has changed in their lives will be.

- 2) Appropriateness:** the evidence is generated through methods that are justifiable given the nature of the purpose of the assessment

Matching the right methods for data collection and analysis with the purpose of the assessment is key to generating quality and robust evidence. Methods need to be based on a clear understanding of what the assessment aims to do. The methods being used to collect data need to be relevant and generate reliable data; the methods for analysis need to be systematic and support the generation of convincing conclusions, and the sample size needs to be in proportion to the conclusions that are sought and the level of validity needed. Also, the team carrying out the assessment needs to be made up of people with the appropriate skills.

- 3) Triangulation:** the evidence has been generated using a mix of methods, data sources, and perspectives

Talking to different groups of people, using different data collection methodologies, and collecting different types of data from different sources helps manage the risk of possible bias. Getting different stakeholders with different perspectives, such as beneficiaries, partners, and peers, to validate findings and conclusions will provide a vital check on their quality.

- 4) Contribution:** the evidence explores how change happens and the contribution of the intervention and factors outside the intervention in explaining change

Robust evidence of change has to explore the causal links. It has to examine the contribution of an intervention, explore whether the assumptions about what causes what are accurate, and examine the relative influence of other factors. There needs to be a point of comparison to anchor any judgement of change. Also, the assessment needs to be actively looking for the unexpected changes and explaining these in the analysis.

- 5) Transparency:** the evidence discloses the details of the data sources and methods used, the results achieved, and any limitations in the data or conclusions

Transparency and openness underpins robust evidence. It is not possible to assess whether evidence is accurate or representative, identify any gaps or potential bias, or assess the logical validity of the analysis, unless the data and the process used to collect and analyse it are clearly presented, including any problems and limitations.

How to use the checklist

Each of the five principles has four key questions and each question can be answered on scale of 1-4. To aid with the scoring and to ensure consistency, scales have been developed for each question. The scales provide a short description of what practice looks like at each of the four levels. These correspond to: 1) weak evidence, 2) minimum standard of evidence, 3) good standard of evidence, 4) gold standard of evidence.

	1. weak evidence	2. minimum standard	3. good standard	4. gold standard
1a. Are the perspectives of beneficiaries included in the evidence?	No beneficiary perspectives presented	Beneficiary perspectives presented, but not integrated into analysis	Beneficiary perspectives presented and integrated into analysis	Beneficiary perspectives presented and integrated into analysis, and beneficiaries have validated the findings; the evidence is strongly grounded in the voices of the poor

Scores for each of the questions are then added up and an overall score for the principles out of 16 is provided. Depending on the score, the principle is then assigned a colour.

An overall score for the principle of 4 – 6	Evidence is weak in this area
An overall score for the principle of 7 – 10	Evidence meets a minimum standard in this area
An overall score for the principle of 11 - 13	Evidence meets a good standard in this area
An overall score for the principle of 14 – 16	Evidence meets gold standard in this area

The table below offers a worked through example of how the scoring would work for the voice and inclusion principle

Principle	Criteria	1	2	3	4	NA	Comments / evidence
1) Voice and Inclusion	1a. Are the perspectives of beneficiaries included in the evidence?				4		
	1b. Are the perspectives of the most excluded and marginalised groups included in the evidence?				4		
	1c. Are the findings disaggregated according to sex, disability and other relevant social differences?	1					
	1d. Do beneficiaries play an active role in designing the assessment process?		2				
Score for voice and inclusion							11/16

A checklist for assessing the quality of evidence (DRAFT)

Evidence being assessed:		Name of assessor:					
Principle	Criteria	1	2	3	4	N A	Comments / evidence
1) Voice and Inclusion We present beneficiaries' views on the effects of the intervention, and identify who has been affected and how	1a. Are the perspectives of beneficiaries included in the evidence?	1	2	3	4		
	1b. Are the perspectives of the most excluded and marginalised groups included in the evidence?	1	2	3	4		
	1c. Are the findings disaggregated according to sex, disability and other relevant social differences?	1	2	3	4		
	1d. Did beneficiaries play an active role in the assessment process?	1	2	3	4		
Score for voice and inclusion:							0/16
2) Appropriateness We use methods that are justifiable given the nature of the intervention and purpose of the assessment	2a. Are the data collection methods relevant to the purpose of the assessment and do they generate reliable data?	1	2	3	4		
	2b. Is the size and composition of the sample in proportion to the conclusions sought by the assessment?	1	2	3	4		
	2c. Does the team have the skills and characteristics to deliver high quality data collection and analysis?	1	2	3	4		
	2d. Do the methods for analysis unpack the data it in a systematic way and produce convincing conclusions?	1	2	3	4		
Score for appropriateness:							0/16
3) Triangulation We make conclusions about the intervention's effects by using a mix of methods, data sources, and perspectives	3a. Are different data collection methodologies used and different types of data collected?	1	2	3	4		
	3b. Are the perspectives of different stakeholders compared and analysed in establishing if and how change has occurred?	1	2	3	4		
	3c. Are conflicting findings and divergent perspectives presented and explained in the analysis and conclusions?	1	2	3	4		
	3d. Are the findings and conclusions of the assessment shared with and validated by a range of key stakeholders (e.g. beneficiaries, partners, peers)?	1	2	3	4		
Score for triangulation:							0/16
4) Contribution We can show how change happened and explain how we contributed to this	4a. Is a point of comparison used to show that change has happened (e.g. a baseline, a counterfactual, comparison with a similar group)?	1	2	3	4		
	4b. Is the explanation of how the intervention contributes to change explored?	1	2	3	4		
	4c. Are alternative factors (e.g. the contribution of other actors) explored to explain the observed result alongside an intervention's contribution?	1	2	3	4		
	4d. Are unintended and unexpected changes (positive or negative) identified and explained?	1	2	3	4		
Score for contribution:							0/16
5) Transparency We are open about the data sources and methods used, the results achieved, and the strengths and limitations of the evidence	5a. Is the size and composition of the group from which data is collected explained and justified?	1	2	3	4		
	5b. Are the methods used to collect and analyse data and any limitations of the quality of the data and collection methodology explained and justified?	1	2	3	4		
	5c. Is it clear who has collected and analysed the data and is any potential bias they may have explained and justified?	1	2	3	4		
	5d. Is there a clear logical link between the conclusions presented and the data collected?	1	2	3	4		
Score for transparency:							0/16

A checklist for assessing the quality of evidence: a scoring scale (DRAFT)

		1	2	3	4
		Weak evidence	Minimum standard of evidence	Good standard of evidence	Gold standard evidence
1. Voice and Inclusion	1a. Are the perspectives of beneficiaries included in the evidence?	No beneficiary perspectives presented	Beneficiary perspectives presented, but not integrated into analysis	Beneficiary perspectives presented and integrated into analysis	Beneficiary perspectives presented and integrated into analysis, and beneficiaries have validated the findings; the evidence is strongly grounded in the voices of the poor
	1b. Are the perspectives of the most excluded and marginalised groups included in the evidence?	No perspectives from most excluded groups presented	Perspectives from most excluded groups presented, but not integrated into analysis	Perspectives from most excluded groups presented and integrated into analysis	Perspectives from most excluded groups presented and integrated into analysis, and excluded groups have validated the findings; the evidence is strongly grounded in the voices of the most excluded
	1c. Are the findings disaggregated according to sex, disability and other relevant social differences?	No disaggregation of findings by social differences	Findings are disaggregated, but a number of social differences relevant to the intervention are missing	Findings are disaggregated according to all social differences relevant to the intervention	Findings are disaggregated according to all social differences relevant to the intervention, and why these have been chosen has been clearly explained
	1d. Did beneficiaries play an active role in the assessment process?	Beneficiaries had no involvement in the assessment process	Beneficiaries had involvement in <u>one</u> of the following: (1) designing the process (2) analysing the data (3) formulating the conclusions	Beneficiaries had involvement in <u>two</u> of the following: (1) designing the process (2) analysing the data (3) formulating the conclusions	Beneficiaries had involvement in <u>all</u> of the following: (1) designing the process (2) analysing the data (3) formulating the conclusions
2. Appropriateness	2a. Are the data collection methods relevant to the purpose of the assessment and do they generate reliable data?	The methods of data collection are not relevant to the purpose of the assessment and/or the data is unreliable	The methods of data collection are relevant to the purpose of the assessment, but there is uncertainty about the reliability of some of the data	Methods of data collection are relevant to the purpose of the assessment and generate reliable data	Methods of data collection are relevant to the purpose of the assessment and generate highly reliable data; there has been appropriate quality control of the data (e.g. spot checks, training data collectors)
	2b. Is the size and composition of the sample in proportion to the conclusions sought by the assessment?	Conclusions are not in proportion to the size and composition of the sample and lack validity	Conclusions claim no more than the size and composition of the sample allows, but there is uncertainty about their validity	Conclusions are in proportion to the size and composition of the sample and are valid	Conclusions are in proportion to the size and composition of the sample and have a high degree of validity
	2c. Does the team have the skills and characteristics to deliver high quality data collection and analysis?	There are doubts about the skills and/or characteristics of the combined team	The combined team appear to have the necessary skills and characteristics	The combined team have demonstrated the necessary skills and characteristics	The combined team have demonstrated both exceptional skills and the characteristics necessary for the task
	2d. Is the data analysed in a systematic way that leads to convincing conclusions?	The method through which the data is analysed is not clear and the conclusions are not convincing	The data is analysed through a clear method, but not every conclusion is wholly convincing	The data is analysed through a clear and systematic method that produces convincing conclusions in all key areas	The data is analysed through a clear and systematic method that produces convincing conclusions in all key areas; there is a detailed analysis of the implications of the conclusions
3. Triangulation	3a. Are different data collection methodologies used and different types of data collected?	Only one data collection method is used	One data collection method is used with reference made to other existing data	Two or more data collection methods and two or more types of data are used	Two or more complimentary and distinct data collection methods and types of data are used
	3b. Are the perspectives of different stakeholders compared and analysed in establishing if and how change has occurred?	Different stakeholder perspectives have not been presented	Different stakeholder perspectives have been presented, but not analysed	Different stakeholder perspectives have been presented and analysed	All stakeholder perspectives relevant to the intervention have been presented and analysed and how and why they have been selected is explained

	3c. Are conflicting findings and divergent perspectives presented and explained in the analysis and conclusions?	Divergent perspectives or conflicting findings are not presented	Divergent perspectives and conflicting findings are presented	Divergent perspectives and conflicting findings are presented and explored	Divergent perspectives and conflicting findings are presented and explored, and there is an in-depth analysis of their implications for the conclusions
	3d. Are the findings and conclusions of the assessment shared with and validated by a range of key stakeholders (e.g. beneficiaries, partners, peers)?	Findings and conclusions are not shared with stakeholders of the intervention	Findings and conclusions are shared with relevant stakeholders of the intervention, but not validated	Findings and conclusions are shared with and validated by relevant stakeholders of the intervention	Findings and conclusions are shared with and validated by all relevant stakeholders of the intervention and their feedback is included in the evidence. The process is taken seriously and this is reflected in the final evidence
4. Contribution	4a. Is a point of comparison used to show that change has happened (e.g. a baseline, a counterfactual, comparison with a similar group)?	No data is available to use as a point of comparison	Data is available and has been used as a point of comparison.	Data is available and has been used as a point of comparison. A clear justification exists for why this is considered appropriate	Data is available and has been used as a point of comparison. A clear justification exists for why this is considered appropriate. The data provides a relevant and high quality basis for comparison.
	4b. Is the explanation of how the intervention contributes to change explored?	No causal links or assumptions are explored	Causal links between the intervention and outcomes are explored	Causal links between the intervention and outcomes and underlying assumptions are explored	All causal link between the intervention and outcomes and underlying assumptions are explored in depth; the evidence provides a clear picture of whether the theory underpinning the intervention's approach to change is sound
	4c. Are alternative factors (e.g. the contribution of other actors) explored to explain the observed result alongside an intervention's contribution?	Analysis does not mention or explore the contribution of factors outside of the intervention	Analysis makes reference to the possible contribution of other factors outside of the intervention	Analysis explores and analyses the contribution of other factors outside the intervention	Analysis provides a comprehensive and systematic analysis of the relative contribution of other factors outside the intervention
	4d. Are unintended and unexpected changes (positive or negative) identified and explained?	Unintended changes are not explored	Unintended changes are identified	Unintended changes are identified and explained	Unintended changes are identified and explained. The methods used for data collection are designed to deliberately capture them
5. Transparency	5a. Is the size and composition of the group from which data is collected explained and justified?	Size and composition of sample are not explained	Size and composition of sample are explained	Size and composition of sample are explained and justified	Size and composition of sample are explained and justified, and all limitations are disclosed
	5b. Are the methods used to collect and analyse data and any limitations of the quality of the data and collection methodology explained and justified?	Methods for data collection and analysis are inadequately explained	Methods for data collection and analysis are explained	Methods for data collection and analysis are explained and justified	Methods for data collection and analysis are explained and justified, and all limitations are disclosed
	5c. Is it clear who has collected and analysed the data and is any potential bias they may have explained and justified?	Team collecting and analysing data not identified and/or potential biases are not mentioned	Team collecting and analysing data are identified and potential biases mentioned	Team collecting and analysing data are identified and potential biases explained and justified	Team collecting and analysing data are identified and potential biases explained and justified; how potential biases are managed is explained
	5d. Is there a clear logical link between the conclusions presented and the data collected?	Conclusions do not follow from the data collected	Conclusions follow from the data collected	Conclusion follow from the data collected and the steps linking them are clearly explained	Conclusion follow from the data collected and the steps linking them are clearly explained; analysis is transparent about limitations of conclusions