

Evaluation Methods and Processes: Tensions between Expectations, Resources, and Competencies

Manfred Max Bergman | *Apart from epistemological aspects, a number of practical factors related to expectations, resources, and competencies bias the selection and application of methods in the course of evaluation processes and thus the validity, objectivity and generalizability of the evaluation results. The article explores these factors and the various tensions between them, and discusses the challenges evaluators and evaluation commissioners have to face in order to avoid unsustainable truth claims and to be able to put evaluation findings into the right perspective. Negotiations with stakeholders or among peer evaluators are seen as two forms of quality control, which can become guide posts for conducting high-quality as well as context-bound evaluation.*

Contents

- 1 Introduction – The limits of objectivity in evaluation
- 2 Sites of tensions in Evaluations
 - 2.1 Context
 - 2.2 Evaluation Method
 - 2.3 Data Collection
 - 2.4 Data Analysis
- 3 Conclusion

1 Introduction – The limits of objectivity in evaluation

Most evaluators would be concordant with the dominant tenor of texts on evaluation: that evaluation¹ should be either objective or at least as objective as possible (e.g. Beywl 2003; Development Assistance Committee 2001; Evaluation Unit 1997; Independent Evaluation Group 2007; OECD 1991; Sera and Beaudry 2007; Sonnichsen 1999). This criterion is clearly linked to a pre-Popperian epistemology, i.e. the idea that rigorous evaluation may capture aspects of the nature of the object of evaluation in ways that do not depend on any idiosyncratic characteristics of the evaluator or the evaluation context. Evaluation in this sense bypasses the 20th century insights into the limits of scientific discovery in relation to truth and validity as developed by, for instance, Popper (1934; 1963) and Kuhn (1962). The social and related sciences, on which evaluation methods and processes draw, have mostly abandoned or at least put into question such old-fashioned, i.e. Comptian², positivistic positions. Among the many problems associated with the pursuit of objectivity are that (a) evaluations take place within a specific context, which makes objectivity extremely difficult at best, (b) evaluations are limited by the knowledge and skills of the evaluators, and (c) the scope of an evaluation is limited to the resources at the disposal of the evaluators. So while expectations

from evaluation commissioners and evaluators themselves about the outcome of the evaluation tend to be very high, tensions and conflicts may arise due to limitations in skills and resources, as well as the evaluation context that underpins the process of evaluation. Consequently, while each final evaluation report either explicitly or implicitly makes rather strong truth claims about the object of evaluation, it is not surprising that the truth claims need to be interpreted as either partial, contextual, politicized – in short, limited.

2 Sites of tensions in Evaluations

It is difficult to discern whether tensions around the evaluation process create unsustainable truth claims about the evaluation process, its methods, or its results, whether evaluators themselves introduce these claims in order to present themselves as experts with special insights into how to identify objective truths, or whether evaluation commissioners, politicians, the public, and other stakeholders make unrealistic demands on evaluators that, ultimately, cannot be fulfilled. While the status quo is likely a result of an interaction between expectations, resources, and competencies, many tensions arise during and after an evaluation, of which the most dominant are related to expectations, resources, and competencies.

The following figure illustrates conceptually the interrelations between the fields of tensions.

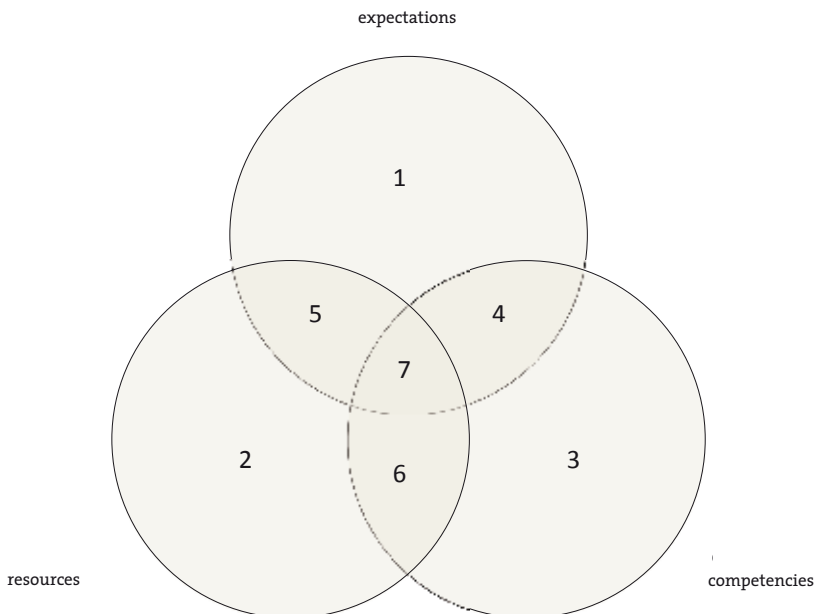


Figure 1: Interrelations between Sites of Tension

Not only are the three domains potential sources of tensions for evaluators and can be described in their own right (represented as areas 1, 2, and 3), but an analysis of tensions in the evaluation process could also examine the relations between tensions arising due to links between expectations and resources (areas 5 and 7), expectations and competencies (areas 4 and 7), and between resources and competencies (areas 6 and 7). For instance, evaluators may not have sufficient time or access to advanced courses in a specific area of evaluation and thus fail to acquire competencies necessary to perform state of the art evaluations. Finally, it may be interesting to assess the extent to which the fields of tension may interact, e.g. how tensions between resources and competencies may depend on expectations (area 7). Continuing the previous example, the necessary update in skills may only become a problem, if any of the stakeholders have expectations about the evaluation that cannot be satisfied with the current skill-level of the evaluator.

Simple and interactive overlap of tensions during the evaluation process can occur at different stages and arise from different sources. The following table illustrates an overview of the sites of tension in relation to expectations, resources, and competencies, on the one hand and the dimension on which these sites of tension may be located, on the other.

	Expectations	Resources	Competencies
Context	11	12	13
Evaluated	21	22	23
Evaluation Method	31	32	33
Data	41	42	45
Analysis	51	52	53
Interpretation	61	62	63
Transfer	71	72	73

Table 1: Sites of Tension and their Dimensionality

Even though this table is quite extensive and thus cannot be covered in detail in this article, it nevertheless must be considered incomplete. There are other sites of tension and there are certainly other types of dimensions on which tensions play themselves out during the evaluation process. I will nevertheless outline some exemplary issues that pertain to the sites of tensions and their dimensionality.

2.1 Context

The evaluation context can include a number of different sub-dimensions, such as professional and institutional norms and ideologies, which are yet again moderated by historical, social, and cultural contexts. For example, the Swiss Federal Office for Public Health at a particular time in history has specific norms on how to submit an evaluation proposal and how to conduct this evaluation (e.g. Evaluation Unit, 1997). This set of norms was different just two decades ago and is likely to change within a decade or two. Other departments/offices in the same country, or offices of public health in other countries, are likely to have different institutional norms on how to apply and conduct an evaluation. It is therefore not only necessary for evaluators to have specific skills relating to applying for and conducting evaluations, but also to be able to discern the different normative climates that exist within a particular sponsoring organisation.

Beyond context as a significant source of tension in itself, it can easily be understood how context relates to the three sites of tension – expectations, competencies, and resources. While there may be high levels of expectations around the results and applicability of the recommendations within an evaluation, a particular institutional context may not allow evaluators sufficient resources to achieve the desired quality. For instance, commissioners may want to generalize the findings from a set of case studies to a larger population but they do not invest the necessary resources that would be required to achieve such a goal. This is often due to the lack of competence of evaluation commissioners, who do not fully understand the limitations of case studies in relation to inference, and this is also due to the lack of competence or disclosure of evaluators, who, for various reasons, may make untenable promises in their evaluation proposals. Other contextual factors that may lead to tensions in relation to expectations, resources, and competences are evaluation habits, i.e. that most evaluators, due to their training and experience, tend to apply only a limited number of evaluation, data collection, and data analysis methods. Similarly, evaluation commissioners may also be convinced that there is only one or a limited number of evaluations, data collection, and data analysis methods suitable for evaluation work. These habitual practices are likely to frame the possibilities and limits of evaluations within the brief of such commissioners.

2.2 Evaluation Method

The textbooks on research methods in the social sciences and, by extension, on evaluation methods, are rather straightforward about the selection of methods: they usually propose to select the method that is most appropriate for an evaluation. Indeed, a number of taxonomies and respective discussions on the pros and

cons of different types of evaluation methods have been produced (e.g. Stufflebeam and Webster 1980; Wholey, Hatry and Newcomer 2004). In theory, one ought to pick the evaluation method most suitable for an evaluation project and the evaluand. This implies a neat one-to-one relationship between a specific evaluation project and a particular evaluation method and evaluand. Unfortunately, the evaluators' professional reality does not behave according to evaluation theory. Instead, there are likely to exist a number of different evaluation methods and, within each method, numerous options open to the evaluator. For instance, to assess the effectiveness of an intervention, most evaluators would select Impact Evaluation as the most appropriate form of evaluation. But even if the broad evaluation method is clear, a number of decisions have to be made, for example in relation to the kind and degree of theory within the evaluation (e.g. Carvalho and White 2004). Evaluators are usually limited by habits, current skills, and available resources. Thus, conflicts can arise due to divergent views on the possibilities and limits of different evaluation methods, how much knowledge and skills are required to conduct a professional evaluation based on a particular evaluation method, how much the evaluand is influenced by different types and degrees of evaluation, and how much of the available resources are required to accomplish just that.

2.3 Data Collection

Numerous considerations need to be raised in relation to the necessity of data for an evaluation and, consequently, numerous sites of tensions can arise. The evaluator or commissioner, who starts with the idea that the data used for an evaluation needs to be unbiased and objective is already in trouble as there is no such thing as objective data. Data are always manufactured in the sense that they are selected, prepared, and interpreted according to the slant that is given to an evaluation project by the stakeholders (Bergman 2008a). For instance, the decision on which document and which part of a document to select for analysis, and how the document or parts thereof should be analysed, is highly idiosyncratic, and will in part depend on expectations, skills, and sometime also on resources. Also rather subjective is the decision about the degree of structure of an interview schedule, whether to conduct group interviews, focus groups, or individual interviews, question and topic order during the interview, etc. While there are no clear guidelines on how to make the process around data collection and data analysis less ambiguous and more objective, any and all decisions around these issues are likely to influence the data and, thus, the evaluation results. Finally, the evaluand may pose limits on what kind of data can be collected from a pragmatic or ethical perspective.

A particular problem in connection to data collection methods relates to the inferences that many evaluation commissioners and even evaluators aim for. In some instances, it is possible to conduct a census so no generalisations from a sample to a population are necessary as the results of the census already represent the qualities of the population. This often is the case when the potential data pool is rather limited and fully accessible to the evaluators. However, in other instances, too many possible interviewees, documents, schools, patients, etc. may exist. While most evaluations do not have the funds to conduct a proper representative study, many commissioners and other stakeholders expect nevertheless that the evaluation results can be expanded beyond the limits of the sample, with which the evaluators worked. Unfortunately, at least from a scientific perspective, this does not work as there are strict sampling rules that govern the possibilities and limits of making population inferences from a sample. Thus, conflicts may arise because of the commissioners' expectations about generalizability of the evaluation findings, which cannot be assured due to a relatively small sample size and, more importantly, an inadequate sampling strategy by the evaluators, as well as a limited budget that does not allow for a more appropriate, systematic data collection relevant to the evaluand.

2.4 Data Analysis

While in most evaluations, data collection is usually limited by time, access, and funding, data analysis is usually limited by skills and habits of evaluators, as well as the expertise and expectations of the evaluation commissioners. Conflicts arise here as well in relation to expectations and resources, or in relation to skills and expectations between these two stakeholders.

A particular problem in relation to data analysis and interpretation of findings rests in the current confusion about the possibilities and limits of qualitative and quantitative methods. The current status quo proposes that qualitative methods are fundamentally different from quantitative methods on ontological, epistemological, and axiological grounds. Many authors speak about the incompatibility thesis between qualitative and quantitative methods, as well as a paradigm war between them (e.g. Denzin and Lincoln 2005; Tashakkori and Teddlie 2003). Thus, it is believed that interviews, for example, are representatives of qualitative research techniques and, thus, must subscribe to a constructivist perspective, while survey research is thought to be a representative of quantitative methods and, thus, ought to subscribe to the (post-)positivist perspective. Following the dominant literature, many evaluators embed their qualitatively oriented evaluation within a constructivist framework, although their evaluation approach, in-

terpretation, and recommendations are everything but constructivist in their nature. Thus, it is quite common to find tremendous inconsistencies between evaluators' theoretical background of their evaluations and their methodological approach on the one hand, and the motivation, interpretations of results, and recommendations, on the other hand. The confusions in the research literature arise due in part to the confusion between data collection methods and data analysis methods, and due to the fact that ontological and epistemological positions are quite unconnected to data collection and analysis (e.g. Bergman 2008b).

3 Conclusion

The main sites of tension in evaluation revolve around expectations, resources, and competences. These three sites can be impinging differently on evaluators, evaluation commissioners, as well as the evaluand. Frequently, the combination or interaction between these sites of tension are experienced differently across different stakeholders. Finally, there are a number of dimensions before, during, and after the evaluation process, within which the sites of tensions play their role. Given the limited space at my disposal, I have focused on only a limited number of combinations in relation to stakeholders (e.g. evaluator and evaluation commissioner), dimensions (e.g. context, data collection, and data analysis), and sites of tensions (expectations, resources, and competence). Whether these tensions play themselves out with regard to resources, expectations among and between evaluators and commissioners, or whether these tensions arise due to inconsistencies in the evaluation between aims, theory, data, analysis, interpretation, and recommendations, etc., it should be recognized that truth claims in relation to objectivity and validity are difficult to make, given state of knowledge and sophistication in relation to the social and related research methods and methodology. So if not objectivity or validity in the classical sense, what else can be the guiding principle to reduce and even eliminate some of these tensions?

Popper proposed directly, Kuhn indirectly, and most pragmatists since that knowledge production is not only a systematic and scientific endeavour, but also a negotiation process among peers. While contemporary social thought often misinterprets this to mean that scientific knowledge and, by extension, evaluation results, are arbitrarily negotiated between stakeholders, thus imposing a constructivist paradigm on these thinkers, they never saw themselves as constructivists. Both Popper and Kuhn insisted that there is such a thing as scientific knowledge and its systematic production, which is fundamentally different from lay peoples' negotiation about meaning. Similarly, we can argue that while negotiations about truth and adequate process has to take place in evaluation as well, the ne-

gotiation partners certainly do not have the liberty to bring to the negotiation table arbitrary ideas. Instead, there are two levels of negotiation and mediation in evaluation. First, negotiations take place at different points during the evaluation process between evaluators and commissioners of evaluations. Sometimes, depending on the nature of the evaluation, the evaluand becomes an important partner in this negotiation process. Second, evaluators also must negotiate with their field and their evaluation peers, which set and negotiate rules and standards. Evaluation training, conferences, and meta-evaluation are three examples of sites at which evaluators not only refine their skills, but also find out and adjust what the current practice is in the evaluation environment, within which they operate. In sum, while classical ideas around objectivity and validity are no longer the standard on which evaluators can measure their efforts, they certainly need to enter into in-depth negotiations with other stakeholders of the evaluation in order to explore expectations, limits, and required skills for the evaluation, as well as with other peer and mentor evaluators, who will assert and impose the standards of their field within a particular place and time. While not as ambitious as objectivity, evaluators are nevertheless guided by these two forms of quality control. A wise evaluator will know how to deal with these tensions such that they become guide posts for conducting high-quality as well as context-bound evaluation.

Manfred Max Bergman, Prof. Dr., Chair in Methodology and Political Sociology, University of Basel, max.bergman@unibas.ch

Footnotes

- 1 While this paper explicitly deals with evaluation methods and processes, many of the arguments made herein could be applied to monitoring and auditing as well.
- 2 Comte is considered the father of positivism. In the first half of the 19th century, he developed the following ideas: that sociology (and the social sciences) should model itself after the natural sciences ("naturalism"); that only observable phenomena offer valid information ("phenomenalism"), things should be reduced to smaller parts to study them appropriately ("atomism"); that science ought to identify general laws ("scientific laws"); and that interpretation and values are of no value in the identification of facts ("axiology"/values and facts). Most researchers no longer subscribe to these tenets as empirical research has never been able to fulfill the expectations and promises of classical positivism. Nowadays, most positivists are actually post-positivists, i.e. a group of researchers who sub-

scribe to the view that a good theory needs to be falsifiable and that a hypothesis can never be proven but that empirical evidence can be used to either reject or fail to reject it (e.g. Popper 1934).

References

- Bergman, Manfred Max (2008a). Combining different types of data for quantitative analysis. In J. Brannen, P. Alasuutari and L. Bickman (Eds.), *Handbook of Social Research Methods*. Thousand Oaks, CA: Sage.
- Bergman, Manfred Max (2008b). *Advances in Mixed Methods Research*. Thousand Oaks, CA: Sage.
- Beywl, Wolfgang (2003). *Selected Comments to the Standards for Evaluation of the German Evaluation Society*. Cologne: DeGEval.
- Carvalho, S. and White, H. (2004). Theory-Based Evaluation: The Case of Social Funds. *American Journal of Evaluation*, 25, 2, 141-160.

- Denzin, N. and Lincoln, Y. (2005). *The SAGE Handbook of Qualitative Research* (3rd ed.). Thousand Oaks, CA: Sage.
- Development Assistance Committee (2001). *Glossary of Evaluation and Results Based Management Terms*. Organisation for Economic Co-Operation and Development.
- Evaluation Unit (1997). *Guidelines for Health Programmes and Project Evaluation Planning*. Bern: Swiss Federal Office of Public Health.
- Independent Evaluation Group (2007). *Sourcebook for Evaluating Global and Regional Partnership Programs: Indicative Principles and Standards*. Washington DC: IEG World Bank.
- Kuhn, T. (1962): *The Structure of Scientific Revolutions*. Chicago: University of Chicago Press.
- OECD (1991). *Principles for Evaluation of Development Assistance*. Paris: OECD.
- Popper, Karl R. (1934). *Logik der Forschung*. Vienna: Springer.
- Popper, Karl R. (1963). *Conjectures and Refutations: The Growth of Scientific Knowledge*. London: Routledge.
- Sera, Y. and Beaudry, S. (2007). *Monitoring and Evaluation*. Social Development Department/World Bank.
- Sonnichsen, R.C. (1999). *Building Evaluation Capacity within Organizations*. In R. Boyle and D. Lemaire (Eds.), *Building Effective Evaluation Capacity: Lessons from Practice*. New Brunswick, NJ: Transaction.
- Stufflebeam, Daniel L. and Webster, W. J. (1980). *An analysis of alternative approaches to evaluation*. *Educational Evaluation and Policy Analysis*. 2(3), 5-19.
- Tashakkori, A. and Teddlie, C. (2003). *Handbook of Mixed Methods in Social and Behavioral Research*. Thousand Oaks, CA: Sage.
- Wholey, Joseph S., Hatry, H.P. and Newcomer, K.E. (2004). *Handbook of Practical Program Evaluation*. San Francisco: Jossey Bass.

Zusammenfassung

Folgt man Wissenschaftstheoretikern wie Popper und Kuhn, hängen die Ergebnisse einer Evaluation immer auch von der Person des Evaluators ab sowie vom Kontext, in dem eine Evaluation stattfindet. Tatsächlich haben nicht nur diese Begleitumstände, sondern auch das begrenzte Knowhow der Evaluatoren sowie beschränkte Ressourcen einen Einfluss auf die Resultate, die beim Streben nach grösstmöglicher Objektivität anfallen. Aus den oftmals hohen Ansprüchen der Auftraggeber einerseits und diesen drei Faktoren – Erwartungen, zur Verfügung stehende Ressourcen und Kompetenzen der Evaluatorinnen und Evaluatoren – andererseits können Konflikte entstehen. Beispielsweise kann ein Auftraggeber fordern, dass die Schlussfolgerungen einer Evaluation stärker generalisiert werden, als dies aus wissenschaftlicher Sicht angezeigt ist. Es gilt somit, auftretende Spannungen zu verringern oder gar zu vermeiden. Dies kann erstens geschehen, indem Evaluator und Auftraggeber an verschiedenen Etappen im Evaluationsprozess Erwartungen, Ressourcen und die benötigten methodischen Ansätze verhandeln; zweitens, indem Evaluatorinnen und Evaluatoren innerhalb der wissenschaftlichen Gemeinde die jeweils geeignetsten Methoden und Regeln erörtern.