

Introduction to Part Three: Evaluation Strategies

Günter Albrecht

Few fields of research in the social sciences have developed so dramatically over the last three decades as evaluation research. One of the reasons for this has probably been that for many years, the major Western societies were firmly committed to reform. This was in line with the claims of the social sciences to provide findings that were of use to society and led to a wave of societal experiments inspired by the social sciences. The scientization of our lives was particularly clearly revealed in that these experiments or reforms were generally not carried out uncritically and without being tested but, in turn, became the object of scientific study. Although the general rules of scientific investigation naturally also apply to this kind of study, it gradually became apparent that the nature of the subject matter of this special type of research simultaneously placed such specific demands on the design and implementation of studies and the presentation of their results that an autonomous methodology developed: namely, evaluation research.

Its characteristics can be determined by stating that its necessary goal is to arrive at a reliable and valid evaluation of the findings despite the fact that the measures subjected to evaluation generally have to be operationalized in a form that only has a limited correspondence to ideal experimental conditions. This is particularly important because it concerns not only the abstract confirmation or rejection of a hypothesis but also the application of finite material and mental resources, so that there is nearly always controversy on the legitimacy, the effectiveness, and the efficiency of these measures.

The results of evaluation research are of interest not only to scientists themselves but also to a range of groups external to science who could be affected in one way or another by the outcome of evaluation research. For this reason, scientific findings from this type of research are not accepted unconditionally; it is more the case that attempts are made to protect vested interests by pointing out actual or supposed errors and methodological inadequacies of evaluation research. These errors can particularly result from the researcher's inability to set up the ideal case of complete experimental conditions. One cardinal problem, for example, is that ethical, legal, organizational, or pragmatic

reasons make it almost impossible to guarantee a strict random assignment of subjects to experimental or control groups. Thus, one absolutely essential prerequisite for a definitive causal attribution of results to treatment is inapplicable. Another typical problem is that, because of costs, for example, evaluation can generally only be carried out for specific regional/local contexts, so that we are faced with the very important question of external validity.

From the long list of other methodological problems, only a few will be mentioned here: problems of standardizing, controlling, and monitoring treatments; the selection of appropriate measurement intervals in order to clearly assign changes in the subjects to the treatment; and so forth.

As most evaluation programs are more or less affected by the methodological problems briefly mentioned above, it is necessary to develop a specific methodology that will permit the most reliable evaluation of the treatment — despite all these potential sources of error. This is the goal of evaluation research methodology. It has developed to such an extent in recent decades that a trend toward the canonization of knowledge can now be seen.

These attempts at canonization — as helpful as they may well be — nonetheless only provide some guidelines. In each concrete case, evaluation researchers are faced with the specific conditions to which they have to apply their methodological arsenal. Before planning a new study, which can cost much time and money, as well as sometimes placing a tremendous strain on the subjects, researchers should ask themselves what knowledge on the practice to be evaluated is actually already available. Due to the increasing scientization of all spheres of life, it should be increasingly rare that no useful research work can be found. It seems to be becoming more typical for certain areas of practice that there is a great amount of research. Nonetheless, this abundance does not provide a clear picture on either an initial or even a second inspection.

In recent years, it has become possible to counter this difficult situation with the new procedure of meta-analysis. By applying this method, one can hope to achieve a systematization of the available knowledge, the establishment of a hard core of knowledge that underlies the complex total picture, as well as indications on theoretical misinterpretations and methodological inadequacies. This method with all its advantages and failings is the object of the first of the three sections on "Evaluation Strategies."

The first article on meta-analysis from Cook succinctly reviews the strengths and limitations of meta-analysis for some specific research questions. Along-

side issues in practical research, this article particularly deals with the underlying problems in the theory of science. In contrast, Cooper's article deals with the more pragmatic aspects of meta-analysis. These are presented so clearly that they will hopefully attract the interest of the many social scientists who have previously had little knowledge about this branch of research.

Hedges' article provides a fundamental and elaborated study on the statistical problems in meta-analysis. These have been investigated very thoroughly in recent years and highly sophisticated solutions are now available. Hedges works out how the particularly difficult details of a meta-analysis can be dealt with. Wittmann's presentation is on a similarly refined statistical and methodological level. He explains why it is so important for evaluation research to consider Brunswik-symmetry if it wants to avoid continuously producing null effects purely as a methodological artifact.

The article from Lösel documents the possible applications and methodological problems of meta-analysis for the example of the relations between family characteristics and behavior disorders in children. These applications and problems have unfortunately received little previous attention in Germany. Lösel is one of the few German scientists who, like Wittmann and Matt, can look back over several years of research experience with meta-analysis. Matt also reports on such experiences. In an interesting way, he uses data on psychotherapy research to show how fruitful meta-analysis can be for subjecting the results aimed for in evaluation research to a secondary analysis.

The second section of Part 3 presents examples from evaluation research that are intended to clarify ethical, political, legal, and methodological-statistical problems. The first article from Treumann uses a single study to show how an overdrawn analysis strategy that ignores certain critical points can lead to serious false conclusions. As it can be assumed that many authors intentionally, though most of them certainly unintentionally, make comparable mistakes, this use of a controversial evaluation problem to provide a detailed demonstration of the dangers that lurk behind an insufficiently critical use of certain methods should be of great assistance.

In contrast, the other examples of research illustrate the particular problems of evaluation research in specific areas of application and for specific types of evaluation. The article from Mager and Hetzel uses the German cardiac prevention study to illustrate the methods used in a process evaluation. The major goal is not to study the effects of specific measures of prevention but to draw attention to which sectors of the population use which means to gain

access to the measures of prevention, so that timely alterations can be planned strategically to improve access to the program.

The article from Häußler-Sczepan addresses the particularly controversial question of the implementation of the new abortion law and its accompanying measures in Germany. The problems that face empirical evaluation research on such a strongly politicized and taboo issue are impressively worked out. At the same time, it becomes clear that even when dealing with such a difficult issue, results can be obtained that are useful for further scientific and political work.

Jörg Albrecht's study is an interesting example of an evaluation study that is not committed to the experimental or quasi-experimental approach but uses the methods of qualitative research to address a very interesting issue in legal sociology. The findings are impressive in that they not only show that a specific form of implementing environmental measures, namely through criminal law, is secondary to environmental protection through administrative law but they also explain why this is so.

Egg's study represents an attempt to use process-produced data to evaluate the policy of treating drug users instead of punishing them. It is demonstrated that even comparatively small resources can provide relevant findings in evaluation research. However, the study also clearly shows that some aspects of the modern laws protecting information on the individual in Germany raise serious barriers for both science and practice.

The article from Kinkel and Josef attempts a kind of research synthesis of diverse studies on the relation between the publicity given to spectacular suicides in the mass media and the occurrence of suicidal behavior in the population following such publicity. This classical issue for communication theory and imitation learning is studied in a way that results in a good example of evaluation research and makes it clear that valuable literature syntheses are also possible without the completely formalistic procedures of meta-analysis, particularly when the data are not based on experimental research designs.

Finally, the article from Günter Albrecht attempts to summarize the major methodological problems in evaluation research and also to break down some redundant positions in the methodological discussion. This is shown to be the correct position by the fact alone that — as the argument demonstrates — flawless designs cannot be selected in the first place because the optimization of one of the four central quality characteristics in evaluation research is always performed to the detriment of one of the other three criteria.

The third section of Part 3 deals with the general importance of longitudinal studies for evaluating research on prevention and intervention or for extending or generating social scientific theories. The reason why this methodological question is so important is that longitudinal studies generally require large financial resources and because they – depending on the specific design – require relatively long or even very long time intervals. As they are also subject to a great number of methodological problems, including, for example, errors in recall in retrospective longitudinal studies, problems of sample mortality, effects of repeated measurement, and the question of generalizability over time, it is questionable whether they are really able to provide the great advances in knowledge that are frequently expected of them. As is nearly always the case, it can also be seen here that the issues are more complex. The articles in this section should emphasize this point.

The first article from Michael Gottfredson and Travis Hirschi provides a very critical discussion of the strong preference for longitudinal studies that can be observed within recent criminological research. He attempts to show that their application to central questions within criminology is neither absolutely necessary nor particularly productive. This certainly applies, with some restrictions, to the testing of theories on the causes of criminality. However, whether longitudinal analysis should be dropped for the evaluation of measures of prevention and intervention and for the analysis of criminal careers is without doubt another issue. Nevertheless, it should certainly be stressed that Gottfredson and Hirschi emphatically show that longitudinal studies are no "cure-all" but, in certain respects, are even markedly inferior to cross-sectional analyses.

Hermann's article presents an interesting attempt to study the problems of recidivism among criminals with a data set that covers considerable time intervals. This study makes it clear that it is impossible to renounce longitudinal data for this topic as, naturally, the cross-sectional data in no way explain the course of recidivism; recidivism is a process with a strong intrinsic momentum.

The brief study from Reuband uses longitudinal data to investigate the problem of the temporal stability of individual attitudes toward criminality and deviance. This problem is particularly important for the question on how attitudes relate causally to long-term behavior patterns.

The article from J. Junger-Tas reports on the temporal course of the effects of a very simple but theoretically well-founded school prevention program. The promising findings of this program and the methods selected make it clear

that careful monitoring in the educational system provides a good basis for effective crime prevention.

Huinink's very complex program uses the example of a large German study to show that it is highly possible to obtain information on important aspects of the relation between social change and individual developments by using retrospective data that meet high methodological criteria.

The same result is obtained with another approach, the panel study, in Strohmeier's work. This documents the strength of specific follow-up surveys for the reconstruction of the dynamics of intimate social relationships.

Naturally, the various examples of different evaluation strategies and the articles on the methodological problems of evaluation research cannot claim to be complete, and we do not even claim to have made a representative presentation of the various paradigms of evaluation research. Even so, these articles should clarify that this research represents a challenge to the scientist and the practitioner of prevention and intervention, and that this challenge can only be met by combining rigorous methodological thinking, skillful ability, and methodological creativity. If this is achieved, then such research represents an important source of information for society and an ideal laboratory for the social scientist.