

BERICHT
aus dem
PSYCHOLOGISCHEN INSTITUT
DER UNIVERSITÄT HEIDELBERG

Joachim Schahn¹ and Gerd Bohner²

**Evaluation Research in Environmental Sociology
and in Environmental Psychology:³ Methodological
Aspects⁴**

November 2002
Diskussionspapier Nr. 85

Inhaltsverzeichnis

Abstract	3
I. Classification of the field	4
II. Topic, objectives, and methods of evaluation research	6
1. Definitions and preliminary remarks	6
2. Experimental design: Experiment or quasi-experiment?	7
2.a Experimental designs	10
2.b Quasi-experimental designs	14
3. Methods of data gathering	19
3.a Reactivity of assessment methods	20
3.b On the problem of socially desirable answers	22
3.c Individual data versus aggregate data	25
4. Tabular summary	27
III. Outlook: Future development of methods and institutional embodiment	28
1. Development of methods	28
2. Institutional embodiment	30
IV. References	32
Institutsberichte des Psychologischen Instituts der Universität Heidelberg	i

¹ Institute for Psychology, University of Heidelberg, Germany

² Institute for Psychology, University of Bielefeld, Germany

³ The authors wish to thank Hans-Peter Erb and an anonymous reviewer for comments on a previous draft of this chapter, and Frank Siebler for his help with the translation from German.

⁴ This paper has previously been released in German language: Schahn, J. & Bohner, G. (1996). Methodische Aspekte sozialwissenschaftlicher Evaluationsforschung im Umweltbereich. In A. Diekmann und C. C. Jaeger (Hrsg.), *Umweltsoziologie, Sonderheft 36/1996 der Kölner Zeitschrift für Soziologie und Sozialpsychologie*, (S. 548-570). Opladen: Westdeutscher Verlag.

Abstract

From a methodological point of view, the authors address social science research dealing with the evaluation of programs in the area of environmental protection. The relevant concepts are defined, and issues of research design and data gathering are discussed. Specifically, experimental and quasi-experimental designs are compared. The authors further discuss problems such as selection effects, treatment contamination, representativeness, reactivity of measurement, influences of social desirability on responses, and the pros and cons of individual versus aggregated data. In an outlook on possible future developments, the aspects of methodology and of institutionalizing evaluation research are addressed.

Key Words:

Environmental Attitudes;

Conservation(Environmental-Behavior);

Program-Evaluation;

Experimental Methods;

Empirical Methods

Authors' Note

Dr. Joachim Schahn, Psychologisches Institut der Universität Heidelberg, Hauptstr. 47-51, D-69117 Heidelberg. Email: Joachim.Schahn@Psychologie.Uni-Heidelberg.de.

Prof. Dr. Gerd Bohner, Fakultät für Psychologie und Sportwissenschaft, Abteilung für Psychologie, Postfach 10 01 31, D-33501 Bielefeld, Email: Gerd.Bohner@Uni-Bielefeld.de

I. Classification of the field

- Jacobs, Bailey and Crews (1984) developed and evaluated a program for encouraging recycling among residential neighborhoods. Recyclable material had to be separated from other garbage and had to be handed to a special collection. Among the measures (also called "interventions") were newspaper advertisements and brochures, distributing containers to help residents separate recyclable from nonrecyclable material, and variations of collection day and collection frequency. Five studies were conducted over a 10-month period to determine the effectiveness of these procedures; several hundred participants were included. Only one group of participants (the experimental group) received the interventions; a second group (the control group) was used for comparison. Finally the most effective interventions were combined into a program and were tested again. The cost of the interventions (e.g., staff, material) were compared to the savings (e.g., money for the recyclable materials, saved costs for dumping or incineration) to assess cost-effectiveness. After a 6-month period, recycling behavior was tested again to find out if possible changes in behavior during the program were maintained afterwards.
- Vining and Ebreo (1989) evaluated an existing recycling campaign in a community. They gathered questionnaire data about knowledge, motivation and self-reported recycling behavior from 443 participants at two points in time (1983 and 1986). Vining and Ebreo did not ask the same respondents at both times (which would have been a longitudinal design), but rather used an independent sample of respondents each time (which is called cross-sectional design). The results demonstrated that all evaluated variables increased (= became more pro-environmental) during the program. A control group (e.g., another community without recycling program) was not included in that study.
- Sexton, Johnson and Konakayama (1987) investigated whether monitoring would encourage consumers to shift their energy demand from the peak- to the off-peak period. 480 households were charged time-of-use prices for electricity, which they could monitor with the help of a special timing device; 120 control households maintained the normal rate structure. Electricity consumption over time was measured as the dependent variable. Both

groups were similar to the customers of the local electricity seller with respect to various demographic variables.

- Kastka (1981) was interested in the effectiveness of a noise protection measure at freeways. People living near the freeway were interviewed one year before and one year after the installation of a noise protection wall. At both points in time, the same people were asked (longitudinal design). Also, before and after the installation of the wall the noise level was measured. Two analogous studies were carried out at two different freeways.

These are typical examples for applied research in the field of environmental protection. Although they address different topics and use different methods, which will be of interest in this article, all methods stem from the arsenal of *evaluation research in the social sciences*. Evaluation research was first introduced in the USA in the 1960's (Wittmann 1985; Wulf 1972). In Germany, evaluation research in the field of environmental protection did not begin to play a major role before the 1980's: E.g., Wittmann (1985: 468) identified education, health and the field of work and economy as "main applications" for evaluation research; he did not mention environmental protection in his book. This situation has changed about the mid-1980's: Since that time, the number of German language publications on environmental topics rose significantly (Schahn 1996), an increase that includes work on evaluation research as well (approx. 5 to 10 studies per year); in the USA the number of publications in that category remained constant on average (about 20 per year), but with considerable variation over the years (e.g., only 8 in 1988, but 43 in 1995; 1996 again only 19). Against that trend, publications about classical interventions based on behavior theory have declined during the 1990's for reasons reported elsewhere (Dwyer, Leeming, Cobern, Porter and Jackson 1993; Geller 1990).⁵

⁵ These data were gathered from an analysis of databases PSYndex, PsycLIT and SocioFile carried out in 1995 and updated in 1998.

Research questions in the field of environmental protection form only part of the topics usually summarized under the notion "environmental psychology"; furthermore, they can also be found in sociology, educational sciences and economics. Therefore, for the purpose of this article we use the notion "environmental research in the social sciences" without differentiation between the various disciplines. This research is defined predominantly by its topic; the theories and methods that are used have mostly been developed in other subdisciplines of the mother sciences. There are no specialized methods of environmental evaluation research. So this article summarizes general methodological considerations of evaluation research, and illustrates them with examples from the field of environmental research, esp. intervention studies that try to change environmentally relevant behavior. We conclude with an outlook to probable future methodological developments and also about developments concerning the general topic "environmental protection". We do not discuss the literature on topics of environmental psychology other than environmental protection (e.g., architecture and physical environment; cf. Preiser 1994; Wener 1989). Also, we do not consider possible interventions to encourage environmental protection at the content level (cf. Freudenberg 1989, Gardener and Stern 1996, Nevin 1991, Oskamp 1983, 1995).

II. Topic, objectives, and methods of evaluation research

1. Definitions and preliminary remarks

Evaluation research may be defined as the application of scientific methods in order to analyze and assess the effectiveness of social measures (projects, programs) with respect to particular objectives (Diekmann 1995: 33ff; Wittmann 1985)⁶. These objectives may include answers to questions like, e.g., the effectiveness of a given measure; whether it should be preferred over another, alternative measure; which specific variables are influenced by the measure; and, whether the employment of a measure can be justified given its costs. Further, two roles of evaluation can be distinguished: while the results of *formative evaluation* are used to change

⁶ Beyond this minimal definition, a great number of proposals and perspectives can be found in the literature. For a detailed discussion including the institutional context and potential goal conflicts between politics and science, see Wittmann (1985: chapter 3).

and optimize measures in progress, *summative evaluation* consists of a final evaluation of programs that are already implemented (Scriven 1972). Among the examples mentioned above, the study of recycling by Vining and Ebreo (1989) can be categorized unequivocally as summative evaluation. Jacobs, Bailey, and Crews (1984) conducted a formative evaluation when testing specific interventions to encourage the collection of recyclable material; their test of the combined program, however, was summative. The Kastka (1981) study qualifies as summative evaluation, since its results did not influence the course of noise protection measures. In contrast, Sexton, Johnson and Konakayama (1987) performed a formative evaluation: their results were fed back into the subsequent improvement of energy savings measures.

Hormuth and Katzenstein (1990: 28ff.) subscribe to a more restricted notion of evaluation, using it exclusively in relation to research that investigates the degree of success of already implemented programs. In addition to what we call summative evaluation, these authors define *model studies* as the testing of a program in a small-scale environment, and *intervention studies* as the comparison of different strategies to improve programs. In both model and intervention studies, the measures under investigation are largely controlled by the investigator, whereas in evaluation studies, these measures have been laid down by a third party. For the present purpose, however, we prefer not to exclude model and intervention studies, but to employ a less restricted notion of evaluation. We start by discussing questions of experimental design. Subsequently, we deal with particular details that require consideration when gathering data. After a summary, we present an outlook on possible and desirable future developments of the research domain.

2. Experimental design: Experiment or quasi-experiment?

With respect to methods adequate for evaluation research, perspectives range from pleas for controlled experiments with random assignment as the method of choice (e.g., Cook and Campbell 1979) to a preference of mere qualitative approaches (e.g., Hamilton, McDonald, King, Jenkins, and Parlett 1977; for a discussion, see Wittmann 1985: 180ff.). Within as well as beyond the domain of environmental evaluation, an experimental approach represents the best

8 Evaluation Research: Methodological Aspects

way to test causal hypotheses and to reject alternative explanations (Seligman and Hutton 1981). Environmental research is *dominated by a quantitative approach*. However, fully randomized field experiments are hardly ever used. Instead, a combination of correlative and quasi-experimental methods (see below) is typically employed.

A study by Dickerson, Thibodeau, Aronson, and Miller (1992) is an example of a fully randomized field experiment. These authors investigated the effectiveness of measures encouraging reduced water consumption when taking showers in public swimming pools. Target persons were to be induced to shower for a shorter time and to turn off the water while soaping up. Showering time was employed as the dependent variable. Two interventions with two treatments each were the independent variables: About half the participants were reminded of prior, water-wasting behavior, while the other participants were not reminded. Orthogonally, some participants were asked for a public commitment to save water, while the others did not engage in such a commitment. Thus, the two variables were crossed in a 2x2 factorial experimental design.

Table 1: Results of the studies by Dickerson, Thibodeau, Aronson, and Miller (1992)

		Commitment	
		no	yes
Reminding	no	301.8 (SD = 142.32)	247.7 (SD = 104.05)
	yes	248.3 (SD = 146.07)	220.5 (SD = 100.62)

Note: N = 80. Values represent average showering time (in seconds). Only in treatment condition "reminding + commitment", showering time was significantly lower than in the control group (no reminding / no commitment). Neither reminding nor commitment alone had a significant impact on showering time.

For the evaluation of this experimental design, it is most important to note that participants were assigned *randomly* to one of the four conditions, and that *no confound* of participation in

the experiment and experimental interventions existed. In the Sexton, Johnson, and Konakayama (1987) study, assignment was random as well, whereas Vining and Ebreo (1989) and Kastka (1981) interviewed convenience samples (see below, II.2.b: Quasi-experiments); furthermore, no control groups were available. Kastka was able at least to collect data from the same persons both before and after the installation of a noise protection wall. The situation is more problematic for the Vining and Ebreo study: Here, the data come from different samples and no control group exists. In both cases, changes in variables between measurements cannot be attributed unequivocally to the effectiveness of the interventions under investigation. Regarding the Vining and Ebreo study, for example, awareness of issues about garbage reduction in the USA might have risen generally within the three years between pretest and posttest⁷.

The most problematic study one can think of in terms of interpretability is undoubtedly the *"one-shot case study"*, where only one group is investigated once after an intervention has been implemented (Cook and Campbell 1979). Still, sometimes there seems to be a necessity even for problematic designs like a *"one-shot case study"* because the alternative would be having no data at all. The relative scarcity of controlled experiments can be traced to unavoidable institutional necessities in context conditions that prevent their employment; it is not due to scientifically reasoned decisions. Most applied settings do not permit, for example, random assignment to experimental conditions: Thus, Vining and Ebreo (1989) had to take advantage of a research opportunity that did not allow for a more sophisticated research design. Nevertheless, we start by considering the advantages of randomized field experiments, trying thereby to illuminate restrictions in the interpretation of results of quasi-experimental designs that exist unless additional controls to safeguard against errors are included (see below). Subsequently, we discuss quasi-experimental alternatives for occasions where a randomized experimental design either has to be given up at some stage of the study or even is not possible to begin with because of contextual constraints.

⁷ The problem of confounding variables will be discussed in sections II.2.a and II.2.b.

2.a Experimental designs

The most important defining feature of an experiment is the construction of groups by *random assignment* to specific treatments (= interventions, measures). This creates an initial comparability of groups, so that differential changes in the criteria under investigation can be traced to a causal influence of the treatment, since additional factors should exert their influence on all investigated groups to a similar amount. Often, the *Solomon-4-groups-design* is presented as the ideal experimental design. By orthogonal variation of "treatment versus no treatment" and "pretest versus no pretest", this design allows to rule out a number of concurrent explanations, i.e., effects of pretest-sensitization, maturity and selection, the influence of external events ("history") and regression towards the mean (Cook and Campbell 1979). The considerable effort involved in this design can be reduced essentially by adopting a pretest-posttest design or a randomized posttest-only design with both an experimental and a control group (Rosenthal and Rosnow 1984). In the study by Dickerson, Thibodeau, Aronson, and Miller (1992) mentioned above, a pretest could be omitted because measurement of the dependent variable was conducted by observation, i.e. *non-reactive* (see below): the observed persons were not aware of the fact that they participated in an experiment, and in one of the four experimental groups, no treatments were employed.

Selection effect in studies with repeated measurement:

For applied evaluation research, pretests are highly recommended even with randomized groups, since in temporarily extended studies a *dropout of participants* is to be expected (*selection effect*). With a pretest, it can be decided whether a dropout from the study occurred due to a selection effect in one of the dependent variables. Of course, this problem can occur only in longitudinal studies (i.e., with more than one point of data collection). An example can be found in a study by Schahn (1996), who investigated the effects of separation of recyclable material in garbage collection: A pretest was conducted in both an experimental and a control group. This pretest allowed to investigate possible differences between the two groups before application of the treatment (changes to the garbage collection system) as well as possible confounds between the relevant variables and further participation in the study. To do so, pretest values of persons participating in both pre- and posttest were compared to pretest values of persons who dropped

out before posttest (statistical method: analysis of variance). This comparison revealed practically no differences, indicating that the selection between assessments was not confounded with relevant variables. One of these variables was the amount of self-reported separation of recyclable material. Assume for a moment that persons who separate recyclable material only to a negligible amount had systematically dropped out after introduction of the new collection system, e.g., because they were not interested being participants of a study about recycling. A seemingly positive effect of the treatment would have been observed, which in fact would reflect little more than differential selection between the two assessments.

Selection effect on first participation:

A related problem is the selection effect on first participation. Typically, not all persons contacted will participate in a study, but only a certain percentage. In practice, the rate differs considerably in a range from about 20 to 90 percent. This kind of selection turns out to be problematic if the reason for non-participation is confounded with variables of interest, e.g., if persons who practice more conservation behavior are more likely to participate than persons who practice less. In this case, the participating groups are not representative subsamples of the defined population (e.g., all inhabitants of a community). Alas, this kind of representativeness cannot be deduced from the sample's participation rate, as Wyss (1990) has made plausible. In principle, representativeness should be demonstrated. This is only possible, however, if non-participants' data are available from additional sources (Binder, Sieber, and Angst 1979; Hirst and Goeltz 1985) or can be gathered by follow-up surveys (Lamnek and Trepl 1991). Usually, however, investigators will be interested in data for which there is no other source than their own study (e.g., data on specific attitudes). A comparison of sample data with the defined population's socio-demographic background variables (e.g., in the community where the study takes place) makes sense only if these variables are presumably related to the relevant characteristics. If demographic variables are unrelated to the studied variables, the representativeness of the sample as far as the variable under investigation is concerned cannot be concluded from the sample's socio-demographically representative composition.

A study by Schahn, Erasmy, Trimpin, and Ditschun (1992, 1994) may be used as an example for the exceptional case where reference data are available that permit a judgment of the sample's representativeness with respect to the investigated characteristics. In this research, garbage avoidance was encouraged by information, feedback and incentives. Participants were asked to comply with checks of their trash bins' contents (participation rate: about 25% of the households asked). It turned out that members of the initial sample exhibited more favorable conservation behavior than did the average member of the community under investigation, even before the intervention took place. Sample data about this behavior were gathered in the study, whereas data for the community as a whole were available from the authorities. This finding holds for both the experimental and the control group. Obviously, participants' motivation was quite high. This could have resulted in a higher acceptance of the treatment by members of the sample than by other (less motivated) members of the community who refused participation, and, accordingly, in an overestimation of the intervention's effectiveness. On the other hand, given that an effect could be demonstrated even though participants scored higher than the reference population's average on target variables to begin with, the effect may be considered even more significant. Finally, both tendencies may have cancelled each other out. Without further research, none of these possibilities can be ruled out; replication studies with different samples may avoid the problem. It is important to note, however, that both investigated groups (experimental and control) were comparable and no confound of prior values and treatment was given. So the questionable representativeness of the sample for the community only threatens the generalizability of the experimental effect (external validity), not the comparison of treatment effects between experimental group and control group (internal validity).

If the variables that lead to biased sampling are known, their effect can be held constant by "*matching*", i.e. by stratification of the groups under investigation according to these variables. A second possibility consists of removing or controlling the confound computationally by several *statistical methods* (regressions, analyses of covariance): The problem of self-selection in participation poses a lesser threat to the evaluation's external validity when the biased variables are controlled statistically (Keating 1989). If sampling errors are overlooked, however, the evaluation's conclusiveness may be dramatically reduced.

"Treatment contamination" and constancy of treatment:

Sometimes it happens that a distinction of experimental versus control conditions cannot be maintained until the end of a program, or that changes to the program are introduced that cannot be controlled by the investigators. In these cases, groups are no longer comparable when it comes to posttest. Such a "*treatment contamination*" could have occurred in the study conducted by Schahn, Erasmy, Trimpin, and Ditschun (1992, 1994). Here, a number of interventions to promote the garbage avoidance on one hand and the accuracy of the separation of garbage from recyclable materials as well as the sorting of those materials on the other were tested. Among the intervention measures was a booklet containing detailed information and advice. Since experimental and control group were recruited from different streets of the same community, interested control group members might have acquired a copy of the booklet from experimental group members they knew. In contrast to the selection effect mentioned above, the consequences of that "treatment contamination" – assuming that it occurred – could have changed results only in one direction: differences between groups could have been reduced, and, hence, the treatment could have appeared less effective. Thus, the possible distribution of booklets to control group members does not provide an alternative explanation for the occurrence of the desired effects but might only have explained their non-occurrence. When recognizing this problem before actually conducting the study, one may include in the posttest questionnaire items that address control group members' potential exposure to interventions. The more time has passed since the end of the study, the more difficult it becomes to verify any suspicion that "treatment contamination" may have prevented the occurrence of expected effects.

A further problem arises particularly with complex field intervention studies, where *not all persons in an experimental condition may receive a truly identical treatment*. This may occur, e.g., when treatments are employed that are based on complex social interactions. In order to promote separated collection of recyclable material in a city-sponsored curbside recycling program, Burn (1991) tried to engage participants as "change agents" who should convince their neighbors in personal communications ("block leader approach"). These interactions presumably took quite different courses and could not be directly controlled by the investigator. In cases like this, continuous measurement of both the quantity and the quality of interventions

is recommended (a "manipulation check" could have been conducted by Burn, e.g. by interviewing change agents and participants). In addition to an experimental comparison between groups, internal correlations of measured intensity of treatment and dependent variable could have been calculated. Thereby, the direction of the effects of differential dropouts or contamination of experimental conditions can often be estimated, and the interpretation of differences between groups is facilitated (Hormuth, Fitzgerald, and Cook 1985; Wittmann 1985).

2.b Quasi-experimental designs

Differing baseline values due to missing random assignment:

In many cases, true experiments cannot be conducted, but quasi-experimental designs have to be employed instead. Quasi-experimental designs can be defined as *comparisons between groups that receive different treatments, without the criterion of random assignment being met*. This method bears the risk that the investigated groups differ not only in the treatment, but additionally with regard to a number of characteristics, each of which may suggest an alternative explanation for observed differences. Hence, the research objective should be to investigate those control groups that are similar to treatment groups on the relevant dimensions, and in particular show comparable pretest values for dependent variables. But meaningful interpretations are possible even with non-comparable baseline values, as long as plausible hypotheses on the influence of these differences can be formulated.

The results of a study by Schahn (1996) provide vivid examples for problematic and unproblematic interpretations in longitudinal studies (here with only two points of measurement). Before and after the introduction of garbage separation, several variables related to the separation and prevention of garbage had been assessed via questionnaire for private households in the experimental community (EC, n = 316) and a control community (CC, n = 246). Results for three relevant variables are shown in Figures 1 to 3. The group means for two points of measurement (1990, and 1991 respectively) are depicted; response scales ranged from 1 to 7.

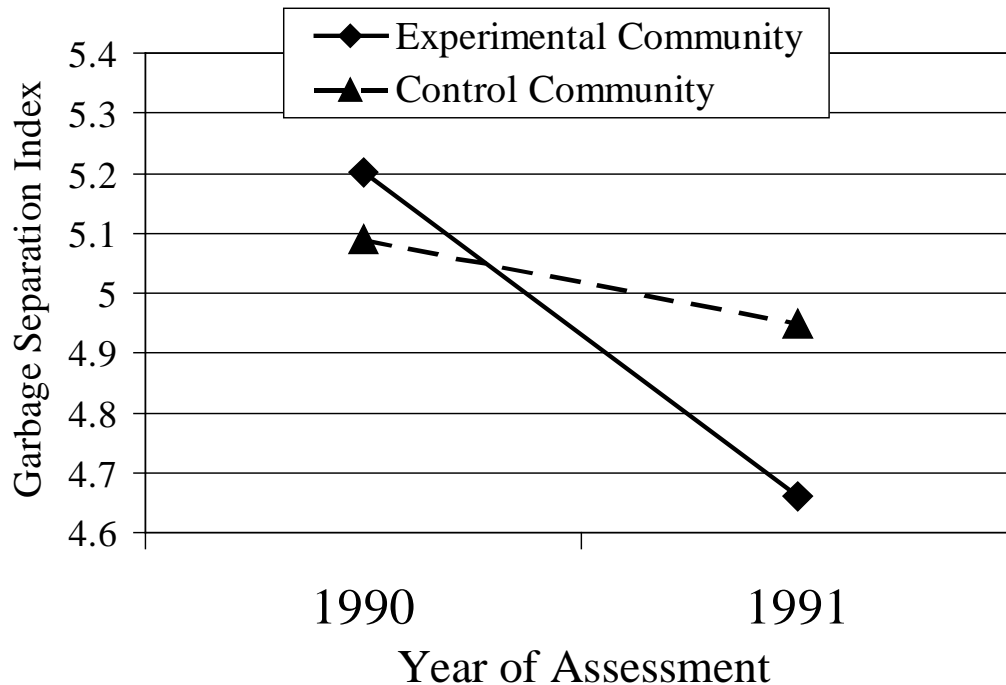


Figure 1: Graphical illustration of the interaction of the Garbage Separation Scale (data from Schahn 1996). For the two assessments in 1990 and 1991, mean values of the "Garbage Separation Index" are depicted separately for each community to illustrate changes.

Figure 1 shows an unproblematic result for a scale that assessed the self-rated use of garbage separation opportunities *in addition* to a newly introduced mandatory community collection for recyclable materials. A decline was predicted. Baseline values of the two communities did not differ significantly at first assessment. An analysis of variance revealed a significant change over time and an interaction: values declined, and the communities differed at second assessment as predicted.

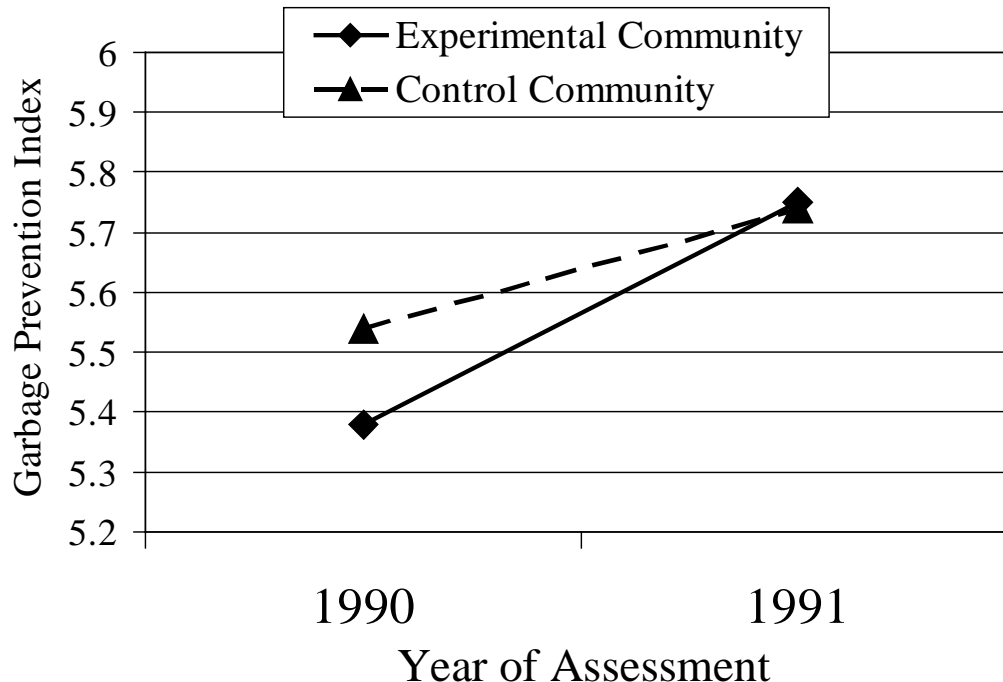


Figure 2: Graphical illustration of the interaction of the Garbage Prevention Scale (data from Schahn 1996). For the two assessments in 1990 and 1991, mean values of the "Garbage Prevention Index" are depicted separately for each community to illustrate changes.

Figure 2 depicts the results for a scale that assessed self-reported garbage-preventing behavior. Here, the pattern is reversed: significant differences at first assessment, improvement of values in both communities, and almost identical values at second assessment (an analysis of variance revealed a significant effect of time and a significant interaction). Possibly, this depicts an effect of the intervention in the experimental community that raised the initially lower value to the higher level of the control, but possibly nothing but regression towards the mean took place. An unequivocal interpretation based on the data given is not possible.

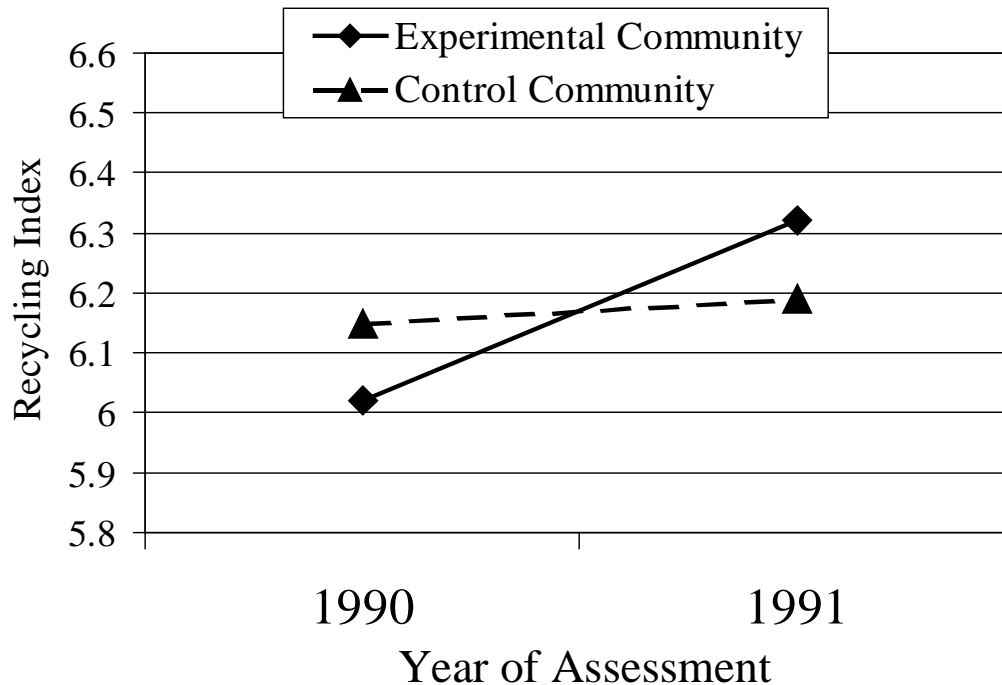


Figure 3: Graphical illustration of the interaction of the Recycling Scale (data from Schahn 1996). For the two assessments in 1990 and 1991, mean values of the "Recycling Index" are depicted separately for each community to illustrate changes.

Figure 3 shows the results for a third scale. The figure illustrates that an interpretation is not necessarily problematic if baseline values differ. On a scale tapping attitudes, behavioral intentions, and self-reported behavior in the field of household garbage, the two samples differed at both assessments, with the experimental community exhibiting a lower value at first assessment, and the control showing a lower value at second assessment. There was no significant temporal development in the control community, whereas the experimental community value increased significantly. Regression towards the mean does not qualify as a sufficient alternative explanation.

In principle, in cases like this, *groups with comparable baseline values* should be identified. For

practical reasons (temporal and financial limitations; in evaluations of already implemented programs; customers' demands or accomplished facts) this will *often not be possible*: for the Schahn (1996) study, only a limited number of communities were available, a situation that enhanced the researchers' tolerance of differing baseline values. A minimal solution to the problem should, at least, consist of the computation of a priori differences and their consideration in the interpretation of effects. More adequate measures to improve the interpretability of research designs with non-comparable control groups are discussed in Hormuth, Fitzgerald, and Cook (1985: 231ff.). One possibility consists of the acquisition of *multiple pretest observations* before implementing a program for all included groups. Thus natural trends in changes of dependent variables can be discovered, and deviations from these trends between the last pretest and the posttest can be interpreted as effects of the program. However, the increased number of pretests not only enhances the overall effort considerably, but may also influence participants' ratings on the behaviors or attitudes under study compared to designs with fewer tests.

A statistical method of correction is "*matching*": in the Schahn (1996) study, e.g., for dependent variables with different baseline values, pairs of persons from each of the two communities with similar values could have been created for analysis. A less desirable consequence would have been, however, that persons with the lowest values in one community and the highest values in the other would not have been included. Additionally, different variables (e.g., attitude scale and garbage prevention scale) might have required different "matching" solutions, which would have introduced biases into the sample composition and reduced inter-variable comparability as well: in the worst case, for each variable, subsamples with slightly different compositions would have to be used.

If no adequate control groups are available at all, *time-series designs including a single study group* may be employed. Two points of measurement as, e.g., in the Kastka (1981) study, can be regarded as a minimal solution. A further strategy to enhance interpretability consists of the attempt to replicate effects. This can be accomplished by employing multiple dependent variables within one study, or by multiple studies with different groups at different times; the

latter solution was chosen by Kastka (1989). Replication is the only possibility whenever the context of investigation prevents the use of control groups as well as time-series designs. Finally, one should attempt to control for known confounding variables statistically, and to design studies in a way that allows subsequent tests of possible concurrent explanations.

The so-called "*multiple baseline design*" is a method of combining several of these objectives in a single study. Here, the same experimental variables are manipulated in at least three different groups; thus, a replication is already built into the design. If the assessments are not conducted simultaneously, but with a slight asynchrony, a number of additional effects can be tested, e.g., selection and "treatment contamination". A study conducted by Schnelle, McNeese, Thomas, Gendrich, and Beagle (1980) provides an example. The authors wanted to test whether feedback about the result of behavioral changes that is spread in a newspaper can serve as an incentive to dispose less waste in the streets, or to even pick up waste ("anti-littering"). The experimental variable consisted of the publication of a figure in the local press that illustrated the number of pieces of waste found in the area under investigation (dependent variable). The study was conducted in three different municipal areas with a temporal displacement. In each of the areas, a considerable decline in the number of pieces of waste resulted. In an additional assessment conducted several weeks after publication of the last article, however, it was found that behavior rates had returned to baseline level. Thus, short- and long-term effects of a treatment need to be distinguished. So-called "*follow-ups*" do not require a different methodological approach; instead, at least one further assessment is conducted after the interventions are withdrawn. In addition to comparisons of pre- and posttest, pretest and follow-up as well as posttest and follow-up can be compared simultaneously or separately. Due to the required effort, however, follow-up studies are typically conducted only within a very narrow time frame (after several weeks at most), or are not conducted at all (for a discussion, see Dwyer, Leeming, Cobern, Porter, and Jackson 1993).

3. Methods of data gathering

Having discussed questions of research design, we now turn to considerations related to the specific procedures of data gathering in evaluation research. Almost the whole arsenal of social

scientific and psychological instruments is in use, ranging from standardized self-report scales and tests to more or less structured interviews, overt and covert behavior observation (e.g., the covert observation of showering time: Dickerson et al. 1992) up to the recording of indirect behavioral traces like consumption of electricity or water as indicated by meters (e.g., Becker and Seligman 1978) or the weight of garbage as well as the quality of separation of recyclable material, as assessed by expert ratings (e.g., Schahn, Erasmy, Trimpin, and Ditschun 1992, 1994). Further, physical measurements may be mentioned, e.g., the assessment of noise level after the introduction of traffic calming measures in a residential neighborhood. To give another example, the evaluation of olfactory emissions cannot rely on physical-chemical measurements alone, but judgments of human "test sniffers" are required in addition (e.g., Kastka 1976; Gellenbeck, Dornbusch, and Gallenkemper 1994). The assessed variables should, of course, be tailored to the specific requirements of the respective study and allow a valid measurement of the target constructs. Accordingly, a sufficient theoretical explication is required. With repeated measurements, the instruments employed need to be sensitive for changes over time (for discussions, see Cook and Campbell, 1979; Wittmann 1985).

We discuss in greater detail two more general aspects that are relevant for the external validity of a study: these are the reactivity of assessment procedures (Webb, Campbell, Schwartz, Sechrest, and Grove 1981) and the problem of social desirability in interview- and questionnaire-based research⁸.

3.a Reactivity of assessment methods

A method of data gathering is to be described as *reactive* if the mere fact of its use triggers a change in the variable that is to be measured (Webb et al. 1981). This may possibly occur with any method that requires participants' cooperation, e.g., in interviews and questionnaires, but also in overt recording of behavior. Effects of the assessment instrument, the situation, the investigator, and the person under investigation need to be distinguished. Since the 1960's, social psychology has investigated these effects, like, e.g., effects of experimenter's

⁸ The level of aggregation of both predictors and dependent variables poses a further problem (for a detailed discussion, see Schahn and Bohner, 1993).

expectancies on the behavior of participants (Rosenthal 1976), the role of subtle hints to the research hypothesis ("demand characteristics") in combination with participants' motivation to behave in a hypothesis-conforming manner (Orne 1962); or the role of additional motivations like evaluation apprehension (Rosenberg 1969) and impression management (Tedeschi, 1981; for an overview, see Diekmann 1995).

These effects should always be considered in social scientific research, since participants are typically informed about the program and its intended results. A distinction needs to be drawn between effects that influence all investigated groups similarly and are, hence, less problematic, as opposed to reactivity effects that influence particular groups differently and, therefore, pose a more serious problem to evaluation. For example, the mere awareness that consumption of electricity or water will be controlled or garbage will be scaled may lead to a more conscious dealing with and reduced use of resources. The relative comparison between groups is not affected by these changes, except in the case of floor or ceiling effects. These might emerge if, e.g. in studies on attitude change, most respondents' attitudes are reflected by the highest scale value, or, in the case of behavioral criteria, if reactivity alone causes minimal consumption that can hardly be further decreased by any treatment. The latter possibility can only be detected by comparison with data from persons who did not participate in a study. For example, in the case of resource consumption, public consumption statistics available from the community may be consulted.

Differential reactivity effects, in contrast, are *more problematic*. Persons who are aware of belonging to a control group might be annoyed or dissatisfied with their role, and consequently cooperate less than members of treatment groups. If effects like this are anticipated, one may take compensating measures in the control group that are suitable for removing differences in motivation but do not influence the dependent variable. For example, a prize for participation might be drawn in a lottery. In social psychological research, a broad range of further measures to eliminate reactivity effects have been proposed and practiced. These measures range from misleading information about the study's true purpose and specific hypotheses on one end of the spectrum to an emphasis on the importance of truthful answers on the other end (see Aronson,

Ellsworth, Carlsmith, and Gonzales 1990).

However, *none of these propositions can eliminate reactivity-induced biases completely*, since participants always have to understand and interpret the meaning of questions and of instructions they receive. Recent studies demonstrate that mere compliance with the conversation rules of everyday life (Grice 1975) causes response biases, even when no particular motivation of a respondent needs to be assumed (Bless, Strack, and Schwarz 1993). In these cases, a remedy is provided by non-reactive assessment methods, like, e.g., the covert observation of behavior, the assessment of behavioral traces (see above), or the study of archival data. Since there are specific disadvantages in non-reactive methods, too, it will usually be advisable to combine them with reactive methods (Webb et al. 1981; and see Bohner 1995).

A good example for the assessment of non-reactive data can be found in a study by Weigel and Newman (1976). In order to validate an attitude questionnaire, each participant that had completed the attitude scale earlier was offered several opportunities to engage in a variety of specific pro-environmental behaviors. First, participants were asked to sign petitions related to environmental concerns. Additionally, respondents were invited to participate in a litter pick-up campaign and to contribute to a recycling program that was continued over several weeks. Participants were led to believe that the requests for these behaviors came from three different organizations and were not related to the study or the questionnaire. For ethical reasons, petitions were indeed submitted to political representatives and the protective activities were in fact conducted.

3.b On the problem of socially desirable answers

Quite a lot of research questions cannot be answered on the basis of behavioral data, whether assessed with reactive or non-reactive methods. These questions are related to behavior in privacy that is difficult or impossible to observe (e.g., taking a shower instead of a bath in order to save resources; turning off the radiator when leaving one's apartment for an extended period of time), or to psychological variables that are, in principle, not directly observable, e.g. attitudes or intentions. These data need to be assessed in self-reports, i.e. in questionnaires or

interviews. Sometimes this method is preferred for reasons of economy even if observation is possible in principle but involves high effort. In this context, concerns about the credibility of collected responses arise. It is often insinuated that participants are aware of the research question (e.g., environmentally relevant behavior), that a norm to behave in an ecologically responsible manner becomes activated, and that *answers are shifted* in this direction: thus, answers are "*distorted*" in a direction that is perceived as "*socially desirable*". This argument is often set forth as a case against the external validity of the results of surveys.

However, the problem is less dramatic than it may appear. First of all, two different cases need to be distinguished: Is the research conducted in order to assess representative population statistics, or to evaluate interventions⁹? In the first case, a distortion would be relatively difficult to deal with, since interest focuses on the absolute values of relevant variables. In the latter case – and only this one is of interest for the purpose of the present contribution – a distortion would not be problematic unless social desirability (SD) affected the groups under investigation to a different extent. Otherwise, interpretation of the data would be impeded only by ceiling or floor effects (see above). *Basically, evaluation studies – just as hypothesis-testing experimental laboratory research – do not require representative samples*, although too specific and, at the same time, homogeneous samples may interfere with the generalizability of intervention effects: The results of an intervention to prevent littering in primary schools cannot easily be transferred to big public events like a carnival parade. But even in this extreme case basic considerations of generalizability from one context to another are possible, at least as far as the relevant variables are known.

Furthermore, it is important to distinguish between cases where behavior is the target variable of an intervention, and cases where attitudes and intentions are to be influenced. With respect to

⁹ Representativeness for, e.g., the population of Germany must not be confused with the representativeness mentioned above of participating persons for a reference sample. In the first case, the "reference sample" would be the whole German population. In the second case, the "reference sample" depends on definition and may consist of, e.g., all inhabitants of a community, or all visitors to public swimming pools (as in the study conducted by Dickerson, Thibodeau, Aronson, and Miller 1992) and so on.

the latter two variables, it is difficult to talk about distortion or bias in a meaningful sense, since there is no objective criterion for the "true" attitude of a person. But even for self-reported behavioral data, at least a *conscious* distortion does not appear too plausible, since respondents are typically not motivated to distort responses in the context of a research study. This has been demonstrated in numerous studies (e.g., Diener 1984; Kury 1983; for a summary, see Mummendey 1987). The *emphasis of anonymity* and the *scientific purpose of the study* are normally adequate measures to reduce SD.

SD can be seen as well as a situational set and as a *comparatively stable personality trait* with different facets (Paulhus 1984, 1986). Attempts to assess SD with separate scales and to control the other variables for SD statistically, or even to eliminate persons with high desirability values from the sample, are based on the latter opinion and are quite questionable (see Mummendey 1987): On the one hand, SD may not affect all persons in the same way and to the same extent; on the other hand, a tendency towards desirability leads some persons to behave, in fact, as prescribed by the respective norm: in this case, SD is not confounded with the relevant variable, but is part of that variable itself (offering help, e.g., is always evaluated in a positive manner). A statistical elimination of SD would then necessarily also eliminate a meaningful part of the variance of the target variable.

Accordingly, questionnaire responses would not reflect conscious or unconscious distortions, but would be quite adequate. In special cases where concern about response distortions is high, specific scaling techniques can be used (e.g., mirroring of items; forced choice among equally desirable alternatives; indirect attitude measures, see Himmelfarb 1993). Typically, however, this is abandoned due to the necessary effort or the lacking availability of appropriate formulations.

These considerations should not be misinterpreted as implying that the assessment of behavioral data is unnecessary and could be replaced by self-reports in questionnaires. Whenever possible, observational data are usually to be preferred over self-reports, if actual behavior is to be assessed (Geller 1981). After all, self-reports in questionnaires refer to past behavior and are not

necessarily good estimators of the probability of desired behavioral *changes*. Alternatively, a questionnaire could be designed to assess a behavioral *intention*. Intentions will, however, not necessarily be transformed into action. The realization of behavioral intentions often depends on additional situational circumstances (Schahn 1993a; for an overview concerning the problem of predicting behavior from attitudes, see Eagly and Chaiken 1993, chapter 4).

3.c Individual data versus aggregate data

All prior considerations referred to the assessment of *individual data* by means of various assessment techniques. This is typically the method of choice in evaluation research, because the researcher's aim is discovering to what extent particular interventions encourage behavioral change or contribute to the persistence or increase of desired behavior in *individuals*. In principle, however, *aggregate data* can be fruitfully employed in evaluation as well¹⁰. For example, to determine whether the introduction of fees for garbage collection that are based on the amount of garbage leads to garbage avoidance, the waste volume (or weight) of several communities that charge such a fee could be compared to the volume of several other communities that charge a rate based on the number of persons in a household. An even better solution would be a before/after comparison across several communities that have recently switched to flexible rates. Since municipal authorities record waste volume anyway, the collection of aggregate data is less expensive than the assessment of individual data. Therefore, several studies in the early 1990's utilized aggregate data. A further advantage of analyses at the aggregate level stems from the fact that the variability of individual decisions needs not be considered. In summative evaluation, this variability is often of little interest and would be treated as error variance anyhow. On the other hand, "aptitude-treatment-interactions" (Cronbach 1975) can only be revealed by assessment of individual data. These interactions allow *sample segmentation in order to tailor interventions* to fit the needs of different population segments (see Geller 1989, Schahn 1995). Such target-group specific intervention strategies may contribute considerably to the improvement of both the magnitude and the persistence of achieved effects.

¹⁰ The concept "aggregate level", as used in this chapter, must not be confused with the concept of "level of aggregation" mentioned before. "Aggregate level" refers to the question whether data stem from individuals or from superordinate units that consist of many individuals, like, e.g., communities. "Level of aggregation" refers to the question whether a given information stems from a single observation, or whether several observations of the same kind (e.g., related items in a questionnaire) have been averaged to enhance reliability.

Another issue in the use of aggregate data is the typical problem of finding a sufficient number of comparable cases, such that, after a division of available cases into experimental and control conditions, meaningful statistical decisions are still possible. This problem does not arise in the example mentioned above, since there are enough communities with different refuse charging systems. In model studies, tests of novel interventions, and/or formative evaluation at the level of aggregate data, however, there is typically only one observable case available, and an analysis of *time-series* data is required. These time-series analyses ("interrupted time series") – whether conducted at the individual or the aggregate level – are contaminated with several (solvable) statistical problems (Campbell 1988: 209ff.). Moreover, as for all kinds of one-case studies, a major disadvantage stems from the limited generalizability of results. Compared to assessments with two points of measurement (before/after) *without control group*, however, this method is considerably more appropriate, since the comparison of two trends allows to control for, at least, regression and maturity-effects. The potential application of time series analyses (for individual as well as aggregate data) is, of course, not restricted to situations where no adequate control group can be identified. Even in common longitudinal studies with both experimental and control group, repeated data assessment before and after interventions provides the opportunity to point out trends, and, thereby, to reject alternative explanations for the obtained effects (or to find explanations for the absence of effects, respectively). Thus, from the viewpoint of a research methodologist, *time-series analyses* may be regarded as the *method of choice*, compared to common designs. The costs are considerably higher research effort and expenditure as well as the requirement to employ specific statistical methods (McCain and McCleary 1979; for references to the relevant literature, see Diekmann 1995: 315). Furthermore, it may not be sufficient to conduct two or three assessments per design phase instead of a single one. The particular strengths of time-series analyses will often show up only with even more frequent assessments. With just two or three assessments per design phase, it is more advisable to average the data in order to enhance their reliability (Schahn and Bohner 1993). These reasons may have contributed to the relative scarcity of evaluation studies that employ time series, at least in the domain of the social sciences¹¹.

¹¹ A prototypical example of a research design tailored to the analysis of time-series is a study by Bullinger (1989). This study does not, however, deal with the evaluation of interventions, but rather investigates the effects of atmosphere pollutants and weather factors on psychological

4. Tabular summary

All the problems mentioned in section II, as well as possible solutions or controls, are summarized in keywords in Table 2.

Table 2: *Table of problems and adequate solutions or controls*

Problem	Solution or control
1. Effect of selection on first participation	- Comparison with reference data - "matching" - statistical control by regression analysis or analysis of covariance
2. Dropout of participants in longitudinal surveys (at least two points of measurement)	- Comparison of relevant pretest variables between persons participating once versus several times
3. Treatment contamination	- Problematic only if treatment effects fail to appear - control group interviews
4. Constancy of treatment	- Checks of quality and quantity of treatment ("manipulation check")
5. Different baseline values of groups under investigation	- Computation of a priori differences and consideration in interpretation (minimal solution) - multiple pretests - attempts to replicate - control of confounding variables (statistically or by "matching") - "multiple baseline design", i.e. delayed repetition of treatment in at least 3 samples
6. Persistence of effects	- Follow-up assessments
7. (Differential) Reactivity of assessment methods	- Non-reactive data assessment - additional measures to enhance motivation in the control group - deception about the study's purpose
8. Socially desirable responses in	- Replace by non-reactive data

state, mental functioning, and physiological arousal of persons who live in areas polluted to a different extent.

surveys	<ul style="list-style-type: none"> - assessment, behavioral traces - emphasis of anonymity and scientific background - specific scaling techniques
9. Simultaneous control of effects due to testing, maturity, selection, external events and regression towards the mean	<ul style="list-style-type: none"> - Solomon-4-groups-design

III. Outlook: Future development of methods and institutional embodiment

It is not the purpose of this section to discuss the future development of behavior change techniques and research strategies at a content level, since an extended body of literature on these topics already exists (e.g., Dwyer et al. 1993; Geller 1990; Leeming, Dwyer, Porter, and Cobern 1993; Oskamp 1983, 1995; Schahn 1995). Instead, we address possible developments of research methods and the institutional embodiment of evaluation research with a stress on developments in Germany, Switzerland and Austria.

1. Development of methods

As already mentioned, environmental research in the social sciences has not developed a specific methodological approach, but applies the social sciences' standard methods to investigate a particular topic. There are no reasons to expect the development of specific methods for the future, at least not independent from or beyond the methodological development of the social sciences more generally. Given the variety of research questions and topics in evaluation research, a true standardization of assessment instruments beyond the already practiced revision of questionnaire scales from time to time does not appear to be possible, although it may be useful for the comparability of results.¹² There is, however, one area that seems to promise an advancement of methods for the years to come: *simulation experiments and computer simulations* as possibilities to reduce research expenses in the preparation of an evaluation. As the examples mentioned above have demonstrated, an evaluation study is always associated with

¹² For a recent approach in the field see Kaiser (1998), who proposed and developed a Rasch-scale for measuring pro-environmental behavior.

considerable expenses, particularly if behavioral data are assessed, and there are some practical problems to abide by the conditions of a true experiment rigidly throughout the study. These problems (see above) may have fatal effects, particularly on time-consuming and non-repeatable longitudinal surveys.

The use of computers provides a possibility to test various interventions "in vitro", and to filter out inefficient interventions even before the actual experiment is conducted in the field. In simulation experiments, participants are introduced to the relevant scenario at a computer (e.g., changes of fees for electricity, water, or public transportation, installation of a new garbage collection system) and their reactions in concrete situations are tested (e.g., willingness to separate different kinds of waste, to refrain from consumption, or choice of transport system). The internal validity of these simulations would be at a maximum, since groups could be randomized and all variables are controlled. The simulation could not only be used to test, in advance, several possible variants of later conducting an evaluation study in the field, but would become an important supplement of the carefully conducted and externally valid field study. Mosler (1993) has used simulation experiments for some time now to investigate ecological dilemma situations where responsible pro-environmental action is encouraged.

Mosler, Gutscher, and Artho (1996) go one step further in their computer simulations of influence processes in social collectives. A novel aspect of their approach is the generation of even the "participants" and their reactions by a computer program. Supposing that the employed algorithms are valid, complex and long-term developments can thus be investigated economically and within a short time frame. Currently, however, the low precision of most theories does not yet allow an exact mathematical implementation, but leaves too many degrees of freedom. This is a limitation of the approach that may prevent its widespread application at least for some time into the future. Furthermore, the design of an appropriate simulation program is very time-consuming, too. Such an effort will only be rational in the case of multiple application possibilities.

2. Institutional embodiment

The examples mentioned in sections I and II are representative for applied research tasks in the environmental domain, since these or similar questions (to reduce consumption of electricity, to enhance quality of garbage separation, and so on) are often posed to persons working in environmentally relevant areas. These persons, in turn, expect assistance by scientific professionals. Most of the described studies were not, however, commissioned work, but basic research (with an orientation towards application), and the interest focused on the effectiveness of theoretical concepts. The results of true commissioned work are seldom published in the scientific literature, either because their implications are not thought to extend beyond the single investigated case, or because the customer is disinterested in a publication. Further, intervention programs are not necessarily motivated by environmental protection concerns; environmental protection may often form just a side effect (for example, if a municipal energy provider attempts to decrease electricity consumption, in order to save the capital expenditure required for a more powerful generating plant). Finally, the evaluation of applied, large-scale interventions with relevance for a whole society has hitherto been the exception in the environmental domain. On the one hand, potential customers may not be aware of the competence that the social sciences and psychology have achieved in this area, or may not ask for external assistance until some of their own attempts have failed. On the other hand, for various reasons, researchers do not actively promote the practical application of basic research results (see Schahn 1993b; Seligman and Hutton 1981). Both aspects result in a waste of both human and financial resources, insofar as adequate measures that have been tested in basic research are not taken, or as municipal councils, authorities, administrations, trade and industry concurrently work at the same problems for which research has already developed solutions or suggestions.

We do not pretend that the social sciences could offer a ready-made solution for any practical problem. What can be offered, in fact, are results and methods that allow, in many cases, to test interventions and to arrive at a scientifically reasoned recommendation. Still missing, however, is the *establishment of professionals offering applications of scientific research* in order to create a link between the universities' basic research and the potential customers. This process

has just started in Germany; few private institutes as well as several dozen self-employed professionals have specialized in the application of social scientific environmental research. In psychology, for example, the section "Umweltpsychologie" (Environmental Psychology) was founded within the "Berufsverband Deutscher Psychologen" (BDP; the German professional association for practically working psychologists) in 1992. The section promotes the development of working opportunities for practitioners in the application of environmental research. The USA, often blazing the trail for similar trends, do not show a greater extent of institutionalization in the area of environmental protection (in the areas of housing and architecture, however, evaluation is a legal requirement in some federal states).

To bring such an applied field into being, the universities are called for as well: The teaching of adequate contents in the various disciplines, as well as simultaneous lobbying are necessary. The relevant contacts should be established by presentations and publications that are acknowledged by potential employers in communities and corporations, trade and industry, and in politics. The social sciences in general, and individual social scientists in particular, need to become known as competent partners in questions of implementation and evaluation of environmental protection measures. Although the traditional definition of academic duties covers mainly research and teaching, professionals in basic research should help bring to life an established field of environmental evaluation research. As long as such extra-academic action ("außer-akademisches Handeln"; Flade and Rohrman 1988: 144) is not regarded as a qualifying feature in the academic context, the task falls to those who have already established themselves at the university. This is because these persons are less challenged to conform to a job profile that emphasizes essentially the quality of research, and to some extent the quality of teaching, but is – at best – disinterested in practical applications that do not result in publications of basic research.

IV. References

- Aronson, Elliot, Phoebe C. Ellsworth, J. Merrill Carlsmith and Marti Hope Gonzales, 1990: *Methods of research in social psychology*. New York: McGraw-Hill.
- Becker, Lawrence J. and Clive Seligman, 1978: Reducing air conditioning waste by signalling it is cool outside. *Personality and Social Psychology Bulletin*, 4: 412-415.
- Binder, Johann, Martin Sieber and Jules Angst, 1979: Verzerrungen bei postalischer Befragung: Das Problem der Nichtantworter [Distortions of data based on questionnaires sent through the mail: The problem of the nonrespondent]. *Zeitschrift für experimentelle and angewandte Psychologie*, 26: 53-71.
- Bless, Herbert, Fritz Strack and Norbert Schwarz, 1993: The informative functions of research procedures: Bias and the logic of conversation. *European Journal of Social Psychology*, 23: 149-165.
- Bohner, Gerd, 1995: Unobtrusive measures. In Antony S. R. Manstead and Miles Hewstone (Eds.), *The Blackwell Encyclopedia of Social Psychology* (pp. 662-664). Oxford: Blackwell.
- Bullinger, Monika, 1989: Psychological effects of air pollution on healthy residents - a time series approach. *Journal of Environmental Psychology*, 9: 103-118.
- Burn, Shawn M., 1991: Social psychology and the stimulation of recycling behaviors: The block-leader approach. *Journal of Applied Social Psychology*, 21: 611-629.
- Campbell, Donald T., 1988: *Methodology and Epistemology for social Science. Selected Papers*. Chicago: University Press.
- Cook, Thomas D. and Donald T. Campbell, 1979: *Quasi- Experimentation: Design and analysis issues for field settings*. Chicago: Rand McNally.
- Cronbach, Lee J., 1975: Beyond the two disciplines of scientific psychology. *American Psychologist*, 30, 116-126.
- Dickerson, Chris Ann, Ruth Thibodeau, Elliot Aronson and Dayna Miller, 1992: Using cognitive dissonance to encourage water conservation. *Journal of Applied Social Psychology*, 22: 841-854.
- Diekmann, Andreas, 1995: *Empirische Sozialforschung. Grundlagen, Methoden, Anwendungen [Empirical social research. Foundations, methods, applications]*. Reinbek: Rowohlt.
- Diener, Edward (1984). Subjective well-being. *Psychological Bulletin*, 235, 542-575.
- Dwyer, William O., Frank C. Leeming, Melissa K. Cobern, Bryan E. Porter, and John Mark Jackson, 1993: Critical review of behavioral interventions to preserve the environment. Research since 1980. *Environment and Behavior*, 25: 275-321.
- Eagly, Alice H. and Shelly Chaiken, 1993: *The psychology of attitudes*. Fort Worth, TX: Harcourt Brace Jovanovich.
- Flade, Antje and Rohrman, Bernd, 1988: Bericht zum Arbeitskreis "Umweltplanung" [Report on the workshop "Environmental Planning"]. In Friedrich Lösel and Helmut Skowronek (Eds.), *Beiträge der politischen Psychologie zu Planungs- und Entscheidungsprozessen* (pp. 142-144). Weinheim: DSV.
- Freudenberg, William R. (1989). Social scientists' contributions to environmental management. *Journal of Social Issues*, 45 (1), 133-152.
- Gardener, Gerald T. and Stern, Paul C. (1996). *Environmental Problems and Human Behavior*. Boston: Allyn and Bacon.

- Gellenbeck, Klaus, Heinz-Josef Dornbusch and Bernhard Gallenkemper, 1994: *Untersuchung zum 2- und 4-wöchigen Restmüllabfuhrhythmus im Landkreis Oldenburg [Survey on the 2- and 4-weekly collection of nonrecyclable garbage in the district of Oldenburg]*. Fachhochschule Münster: Labor für Abfallwirtschaft, Siedlungswasserwirtschaft und Umweltchemie (LASU).
- Geller, E. Scott, 1981: Evaluating energy conservation programs: Is verbal report enough? *Journal of Consumer Research*, 8: 331-335.
- Geller, E. Scott, 1989: Applied behavior analysis and social marketing: An integration for environmental preservation. *Journal of Social Issues*, 45 (1): 17-36.
- Geller, E. Scott, 1990: Behavior analysis and environmental protection: Where have all the flowers gone? *Journal of Applied Behavior Analysis*, 23: 269-273.
- Hamilton, D., B. McDonald, C. King, D. Jenkins and M. Parlett, (Eds.), 1977: *Beyond the numbers game: A reader in educational evaluation*. Berkeley: McCutchan.
- Himmelfarb, Samuel (1993). The measurement of attitudes. In Alice H. Eagly and Shelly Chaiken, *The psychology of attitudes* (pp. 23-87). Fort Worth, TX: Harcourt Brace Jovanovich.
- Hirst, Eric and Richard Goeltz, 1985: Accuracy of self-reports: Energy conservation surveys. *Social Science Journal*, 22: 19-30.
- Hormuth, Stefan E. and Henriette Katzenstein, 1990: *Psychologische Ansätze zur Müllvermeidung und Müllsortierung. Forschungsbericht für das Ministerium für Umwelt Baden-Württemberg [Psychological approaches for avoidance and separation of garbage. Research report for the Ministry of Environment of the state Baden-Württemberg]*. Heidelberg: Psychologisches Institut der Universität.
- Hormuth, Stefan E., Nancy F. Fitzgerald and Thomas D. Cook, 1985: Quasi-experimental methods for community-based research. In Edwin C. Susskind and Donald C. Klein (Eds.), *Community Research* (pp. 206-249). New York: Praeger.
- Jacobs, Harvey E., Jon S. Bailey and James I. Crews, 1984: Development and evaluation of a community based resource recovery program. *Journal of Applied Behavior Analysis*, 17: 127-145.
- Kaiser, Florian G., 1998: A general measure of ecological behavior. *Journal of Applied Social Psychology*, 28: 395-422.
- Kastka, Joachim, 1976: Untersuchungen zur Belästigungswirkung der Umweltbedingungen Verkehrslärm und Industriegerüche [Studies on the annoyance effect due to the environmental factors traffic noise and industrial odors]. In Gerhard Kaminski (Ed.), *Umweltpsychologie* (pp. 187-223). Stuttgart: Klett.
- Kastka, Joachim, 1981: Untersuchungen zur subjektiven Wirksamkeit von Maßnahmen gegen Verkehrslärm und deren Moderation durch nichtakustische Faktoren [Studies on the subjective effectiveness of measures against traffic noise and their moderation by non-acoustic factors]. In Henning Haase and Walter Molt (Eds.), *Handbuch der Angewandten Psychologie, Bd. 3: Markt und Umwelt* (pp. 468-485). Landsberg: Verlag Moderne Industrie.
- Keating, Kenneth M., 1989: Self-selection: Are we beating a dead horse? *Evaluation and Program Planning*, 12: 137-142.
- Kury, Helmut, 1983: Zur Verfälschbarkeit von Persönlichkeitsfragebogen bei jungen Strafgefangenen [Falsification of personality questionnaires by juvenile offenders]. *Zeitschrift für Strafvollzug und Straffälligenhilfe*, 32: 323-332.

- Lamnek, Siegfried and Ralph Trepl, 1991: Die Nichtteilnahme an sozialwissenschaftlichen Befragungen. Analyse der Nachfassaktion bei einer postalischen Repräsentativumfrage [The nonresponse in social science surveys. Analysis of a representative follow-up mail survey]. *Planung und Analyse*, 18: 205-211.
- Leeming, Frank C., William O. Dwyer, Bryan E. Porter and Melissa K. Cobern, 1993: Outcome research in environmental education: A critical review. *Journal of Environmental Education*, 24: 8-21.
- McCain, Leslie J. and Richard McCleary, 1979: The statistical analysis of the simple interrupted time-series quasi-experiment. In Thomas D. Cook and Donald T. Campbell, *Quasi-experimentation* (pp. 233-293). Boston: Houghton Mifflin.
- Mosler, Hans-Joachim, 1993: Self-dissemination of environmentally-responsible behavior: The influence of trust in a commons dilemma game. *Journal of Environmental Psychology*, 13: 111-123.
- Mosler, Hans-Joachim, Heinz Gutscher and Jürg Artho (1996). Kollektive Veränderungen zu umweltverantwortlichem Handeln Handeln [Collective changes towards environmentally responsible acting]. In Ruth Kaufmann-Hayoz and Antonietta Di Giulio (Eds.), *Umweltproblem Mensch. Humanwissenschaftliche Zugänge zu umweltverantwortlichem Handeln* (pp. 237-260). Bern: Haupt.
- Mummendey, Hans D., 1987: *Die Fragebogen-Methode. Grundlagen und Anwendung in Persönlichkeits-, Einstellungs- und Selbstkonzeptforschung [The questionnaire method. Foundations and application in personality, attitude, and self concept research]*. Göttingen: Hogrefe.
- Nevin, John A. (1991). Behavior Analysis and global survival. In W. Ishaq (Ed.), *Human Behavior in Today's World* (pp. 39-49). New York: Praeger.
- Orne, Martin T., 1962: On the social psychology of the psychological experiment: With particular reference to demand characteristics and their implications. *American Psychologist*, 17, 776-783.
- Oskamp, Stuart, 1983: Psychology's role in the conserving society. *Population and Environment: Behavioral and Social Issues*, 6: 255-293.
- Oskamp, Stuart, 1995: Applying social psychology to avoid ecological disaster. *Journal of Social Issues*, 51 (4): 217-238.
- Paulhus, Delroy L., 1984: Two-component models of socially desirable responding. *Journal of Personality and Social Psychology*, 46: 598-609.
- Paulhus, Delroy L., 1986: Self-deception and impression management in test responses. In Alois Angleitner and Jerry S. Wiggins (Eds.), *Personality assessment via questionnaires. Current issues in theory and measurement* (pp. 143-165). Berlin: Springer.
- Preiser, Wolfgang F. E., 1994: Built environment evaluation: Conceptual basis, benefits and uses. *Journal of Architectural Planning and Research*, 11: 91-107.
- Rosenberg, Milton J., 1969: The conditions and consequences of evaluation apprehension. In Robert Rosenthal and Ralph L. Rosnow (Eds.), *Artifact in behavioral research* (pp. 279-349). New York: Academic Press.
- Rosenthal, Robert, 1976: *Experimenter effects in behavioral research* (2nd ed.). New York: Appleton-Century-Crofts.
- Rosenthal, Robert and Ralph L. Rosnow, 1984: *Essentials of behavioral research*. New York: McGraw-Hill.

- Schahn, Joachim, 1993a: Die Kluft zwischen Einstellung und Verhalten beim individuellen Umweltschutz [The gap between attitude and behavior in individual environmental protection]. In Joachim Schahn and Thomas Giesinger (Eds.), *Psychologie für den Umweltschutz* (pp. 29-49). Weinheim: Beltz/PVU.
- Schahn, Joachim, 1993b: Psychologische Beiträge zum Umweltschutz: Forschung und Anwendung [Psychological contributions to environmental protection: Research and application]. In Joachim Schahn and Thomas Giesinger (Eds.), *Psychologie für den Umweltschutz* (pp. 63-75). Weinheim: Beltz/PVU.
- Schahn, Joachim, 1995: Psychologische Forschung zu Mülltrennung und Müllvermeidung: Möglichkeiten und Grenzen psychologischer Ansätze zur Förderung umweltschonenden Verhaltens [Psychological research on garbage separation and garbage avoidance: Opportunities and limits of psychological approaches for promoting pro-environmental behavior]. *Psychologische Rundschau*, 46: 104-114.
- Schahn, Joachim, 1996: *Die Erfassung und Veränderung des Umweltbewußtseins [Assessment and change of environmental awareness]*. Frankfurt/M.: Lang.
- Schahn, Joachim and Gerd Bohner, 1993: Aggregation oder Desaggregation? Einige Bemerkungen zur Debatte um die Ergebnisse von Diekmann und Preisendorfer [Aggregation or disaggregation? Some remarks on the controversy concerning Diekmann and Preisendorfer's research]. *Kölner Zeitschrift für Soziologie und Sozialpsychologie*, 45: 772-777.
- Schahn, Joachim, Petra Erasmy, Andrea Trimpin and Kersten Ditschun, 1992: *Erprobung psychologischer Maßnahmen zur Förderung von Hausmüllvermeidung und Hausmülltrennung [Evaluation of psychological interventions to promote prevention and separation of garbage]*. Forschungsbericht für das Ministerium für Umwelt Baden-Württemberg. Heidelberg: Psychologisches Institut der Universität.
- Schahn, Joachim, Petra Erasmy, Andrea Trimpin and Kersten Ditschun, 1994: *Psychologische Maßnahmen zur Förderung von Hausmüllvermeidung und Hausmülltrennung [Psychological interventions to promote prevention and separation of garbage]*. Bericht aus dem Psychologischen Institut der Universität Heidelberg, Diskussionspapier Nr. 78, Februar 1994. Heidelberg: Psychologisches Institut der Universität.
- Schnelle, John F., M. Patrick McNees, Murphy M. Thomas, John G. Gendrich and Gwen P. Beagle, 1980: Prompting behavior change in the community: Use of mass media techniques. *Environment and Behavior*, 12: 157-166.
- Scriven, Michael, 1972: Die Methodologie der Evaluation [The methodology of evaluation]. In Christoph Wulf (Ed.), *Evaluation: Beschreibung und Bewertung von Unterricht, Curricula und Schulversuchen* (pp. 60-91). München: Piper.
- Seligman, Clive and R. Bruce Hutton, 1981: Evaluating energy conservation programs. *Journal of Social Issues*, 37 (2): 51-72.
- Sexton, Richard J., Nancy B. Johnson and Akira Konakayama, 1987: Consumer response to continuous-display electricity-use monitors in a time-of-use pricing experiment. *Journal of Consumer Research*, 14: 55-62.
- Tedeschi, James T., 1981: *Impression management*. New York: Academic Press.
- Vining, Joanne and Angela Ebreo, 1989: An evaluation of the public response to a community recycling education program. *Society and Natural Resources*, 2: 23-36.

- Webb, Eugene J., Donald T. Campbell, Richard D. Schwartz, Lee Sechrest and Janet Belew Grove, 1981: *Nonreactive measures in the social sciences (2nd ed.)*. Boston, MA: Houghton Mifflin.
- Weigel, Russell H. and Lee S. Newman, 1976: Increasing the attitude-behavior correspondance by broadening the scope of the behavioral measure. *Journal of Personality and Social Psychology*, 33: 793-802.
- Wener, Richard, 1989: Advances in evaluation of the built environment. In Ervin H. Zube and Gary T. Moore (Eds.), *Advances in evaluation of the built environment* (pp. 287-313). New York: Plenum.
- Wittmann, Werner W., 1985: *Evaluationsforschung: Aufgaben, Probleme und Anwendungen [Evaluation research: Objectives, problems, and applications]*. Berlin: Springer.
- Wulf, Christoph (Ed.), 1972: *Evaluation: Beschreibung und Bewertung von Unterricht, Curricula und Schulversuchen [Evaluation: Description and assessment of lessons, curriculae, and school experiments]*. München: Piper.
- Wyss, Werner, 1990: Darf eine heilige Kuh geschlachtet werden? Die 70-Prozent-Marke bei der Ausschöpfung von Random-Umfragen [May a holy cow be slaughtered? The 70-percent mark in completion of random surveys] *Planung und Analyse*, 17: 69-71.

**Institutsberichte des Psychologischen Instituts der Universität Heidelberg
(incl. Historische Reihe)**

Hauptstraße 47-51, D-69117 Heidelberg
Telefon: +49-6221 / 54-7346, Fax: +49-6221 / 54-7734

Preis pro Seite bei Bestellung: € 0,10 (Stand: Januar 2002)

Diskussionspapier Nr. 1:

Groeben, N.: Vom behavioralen zum epistemologischen Subjektmodell: Paradigmawechsel in der Psychologie? (September 1975)

Diskussionspapier Nr. 2:

Möbus, C. & Simons, H.: Zur Fairness psychologischer Intelligenztests gegen Gruppen: Kritik klassischer Konzepte. (Oktober 1975)

Diskussionspapier Nr. 3:

Wottawa, H.: Skalenprobleme bei probabilistischen Meßmodellen. (März 1976)

Diskussionspapier Nr. 4:

Treiber, B. & Petermann, F.: Zur Interaktion von Lernermerkmalen und Lehrmethoden: Rekonstruktion und Normierung des ATI-Forschungsprogramms. (April 1976)

Diskussionspapier Nr. 5:

Möbus, C. & Wallasch, R.: Zur Erfassung von Hirnschädigungen bei Kindern: Nichtlineare Entscheidungsregeln auf der Basis von Veränderungsmessungen. (August 1976)

Diskussionspapier Nr. 6:

Scheele, B. & Groeben, N.: Voraussetzungs- und zielspezifische Anwendung von Konditionierungs- vs. kognitiven Lerntheorien in der klinischen Praxis. (Dezember 1976)

Diskussionspapier Nr. 7:

Möbus, C.: Zur Analyse nichtsymmetrischer Ähnlichkeitsurteile: Ein dimensionales Driftmodell, eine Vergleichshypothese, TVERSKY's Kontrastmodell und seine Fokushypothese. (Juni 1977)

Diskussionspapier Nr. 8:

Simons, H. & Möbus, C.: Veränderung von Berufschancen durch Intelligenztraining. (Juli 1977)

Diskussionspapier Nr. 9:

Braunmühl, C. v. & Grimm, H.: Zur Kommunikationspsychologie: Über Versuche der methodischen Konstitution eines genuin humanwissenschaftlichen Forschungsansatzes zur Entwicklung der Verständigungsfähigkeit. (November 1977)

Diskussionspapier Nr. 10:

Hofer, M.: Entwurf einer Heuristik für eine theoretisch geleitete Lehrer- und Erzieherbildung. (November 1977)

Diskussionspapier Nr. 11:

Scheibler, D. & Schneider, W.: Probleme und Ergebnisse bei der Evaluation von Clusteranalyse-Verfahren. (Juni 1978)

Diskussionspapier Nr. 12:

Scheele, B.: Kognitions- und sprachpsychologische Aspekte der Arzt-Patient-Kommunikation. (September 1978)

- Diskussionspapier Nr. 13:
 Treiber, B. & Schneider, W.: Mehrebenenanalyse sozialstruktureller Bedingungen schulischen Lernens. (Oktober 1978)
- Diskussionspapier Nr. 14:
 Ahrens, H.-J. & Kordy, H.: Möglichkeiten und Grenzen der theoretischen Aussagekraft von multidimensionalen Skalierungen bei der Untersuchung menschlicher Informationsverarbeitung. Teil I: Formale und wissenschaftstheoretische Grundlagen. (März 1979)
- Diskussionspapier Nr. 15:
 Groeben, N.: Entwurf eines Utopieprinzips zur Generierung psychologischer Konstrukte. (Juni 1979)
- Diskussionspapier Nr. 16:
 Weinert, F.E. & Treiber, B.: School Socialization and Cognitive Development. (Juni 1979)
- Diskussionspapier Nr. 17:
 Gundlach, H.: Inventarium der älteren Experimentalapparate im Psychologischen Institut Heidelberg sowie einige historische Bemerkungen. (1978)
- Diskussionspapier Nr. 18:
 Scheele, B. & Groeben, N.: Zur Rekonstruktion von Subjektiven Theorien mittlerer Reichweite. Eine Methodik-Kombination von halbstandardisiertem Interview (einschließlich Konfrontationstechnik) und Dialog-Konsens über die Theorie-Rekonstruktion mittels der Struktur-lege-Technik (SLT). (Dezember 1979)
- Diskussionspapier Nr. 19:
 Gloger-Tippelt, G.: Subjektive Theorien von Frauen über ihre erste Schwangerschaft: Theoretische Konzepte und methodische Möglichkeiten. (Januar 1980)
- Diskussionspapier Nr. 20:
 Kämmerer, A.: Das Konzept 'psychotherapeutische Strategie' am Beispiel des Problemlösens. (Juli 1980)
- Diskussionspapier Nr. 21:
 Scheele, B. (unter Mitarbeit von B. Tuschen und C. Maier): Subjektive Theorien über Ironie – als Heuristik für einen wissenschaftlichen Hypothesenkörper. (August 1980)
- Diskussionspapier Nr. 22:
 Treiber, B.: Erklärung von Förderungseffekten in Schulklassen durch Merkmale subjektiver Unterrichtstheorien ihrer Lehrer. (Oktober 1980)
- Diskussionspapier Nr. 23:
 Röhrle, B. & Kommer, D.: Handlungstheoretische Betrachtungen zur primären Prävention psychischer Störungen. (Februar 1981)
- Diskussionspapier Nr. 24:
 Voigt, F.: Die Entwicklung des Zahlbegriffs. Teil I: Entwicklungslinien des Zahlbegriffs im Vorschulalter: Übersicht über theoretische Probleme und empirische Untersuchungen, mit einer Bibliographie zur Zahlbegriffsentwicklung. Teil II: Entwicklungslinien des Zahlbegriffs im Vorschulalter: Deskriptive Untersuchung des kindlichen Zahlverständnisses und verwandter Konzepte. (April 1981. Teil III: Trainingsstudien zum Erwerb konkreter Operationen (unter besonderer Berücksichtigung von Modellen der Invarianzaufgabe). Teil IV: Die Trainierbarkeit ordinaler und kardinaler Konzepte und ihre Beziehung zum Zahlbegriff. (Juli 1982)
- Diskussionspapier Nr. 25:
 Schneider, G. & Weimer, E.: Aspekte der Kategorisierung städtischer Umwelt – Eine empirische Untersuchung. (Juni 1981)

- Diskussionspapier Nr. 26:
Schneider, W. & Scheibler, D.: Zur Evaluation numerischer Klassifikation: Probleme beim Vergleich von Clusteranalysen. (August 1981)
- Diskussionspapier Nr. 27:
Drinkmann, A. & Groeben, N.: Techniken der Textorganisation zur Verbesserung des Lernens aus Texten: Ein metaanalytischer Überblick. (November 1981)
- Diskussionspapier Nr. 28:
Graumann, C.F.: Theorie und Geschichte. (November 1982, Historische Reihe Nr. 1)
- Diskussionspapier Nr. 29:
Woodward, W.R.: From the Science of Language to Völkerpsychologie: Lotze, Steinthal, Lazarus and Wundt. (November 1982, Historische Reihe Nr. 2)
- Diskussionspapier Nr. 30:
Sommer, J.: Dialogische Forschungsmethoden. (Dezember 1982)
- Diskussionspapier Nr. 31:
Wintermantel, M. & Christmann, U.: Textverarbeitung: Empirische Untersuchung zum Verstehen einer Personbeschreibung. (Januar 1983)
- Diskussionspapier Nr. 32:
Schmalhofer, F.: Text Processing with and without Prior Knowledge: Knowledge- versus Heuristic-Dependent Representations. (Februar 1983)
- Diskussionspapier Nr. 33:
Métraux, A.: Victor de l'Aveyron oder Zum Streit zwischen Kulturalisten und Biologen am Anfang des 19. Jahrhunderts. (Mai 1983, Historische Reihe Nr. 3)
- Diskussionspapier Nr. 34:
Graumann, C.F.: Wundt – Bühler – Mead – Zur Sozialität und Sprachlichkeit menschlichen Handelns. (Mai 1983, Historische Reihe Nr. 4)
- Diskussionspapier Nr. 35:
Gundlach, H.: Folk Psychology and Social Psychology oder Das Los des Ausdrucks 'Völkerpsychologie' in den englischen Übersetzungen der Werke Wundts. (Mai 1983, Historische Reihe Nr. 5)
- Diskussionspapier Nr. 36:
Woodward, W.R.: Hermann Lotze's Concept of Function: Its Kantian Origin and its Impact on Evolutionism in the United States. (Mai 1983, Historische Reihe Nr. 6)
- Diskussionspapier Nr. 37:
Schneider, G.: Reflexivität als Grenzproblem einer kognitiven Psychologie. (August 1983)
- Diskussionspapier Nr. 38:
Geuter, U.: 'Gleichschaltung' von oben? Universitätspolitische Strategien und Verhaltensweisen in der Psychologie während des Nationalsozialismus. (Oktober 1983, Historische Reihe Nr. 11)
- Diskussionspapier Nr. 39:
Kruse, L.: Drehbücher für Verhaltensschauplätze oder: Scripts for Settings. (Dezember 1983)
- Diskussionspapier Nr. 40:
Graumann, C.F.: The individualisation of the social and the desocialisation of the individual – Floyd H. Allport's Contribution to Social Psychology. (Mai 1984, Historische Reihe Nr. 10)
- Diskussionspapier Nr. 41:
Kruse, L. & Graumann, C.F.: Environmental Psychology in Germany. (November 1984)

Diskussionspapier Nr. 42:

Kany, W. & Schneider, G.: Ein linguistisch fundiertes inhaltsanalytisches System zur Erfassung des referentiellen und prädikativen Gehalts verbaler Daten. (Mai 1985)

Diskussionspapier Nr. 43:

Hormuth, S.E.: Methoden für psychologische Forschung im Feld: Erfahrungsstichprobe, Autophotographie und Telefoninterview. (Februar 1985)

Diskussionspapier Nr. 44:

Haerberle, E.J.: Die Anfänge der Sexualwissenschaft in Berlin. (April 1985, Historische Reihe Nr. 12)

Diskussionspapier Nr. 45:

Schmalhofer, F. & Schäfer, I.: Lautes Denken bei der Wahl zwischen benannt und beschrieben dargebotenen Alternativen. (Juni 1985)

Diskussionspapier Nr. 46:

Zielinski, W. & Rott, C.H.: Analyse der Entwicklung des Wortleseprozesses bei erfolgreichen und schwachen Lesern der Grundschule. (Februar 1986)

Diskussionspapier Nr. 47:

Waller, M.: Metasprachliche Entwicklung: Forschungsgegenstand, Schwerpunkte, Desiderate und Perspektiven der empirischen Forschung. (Juli 1986)

Diskussionspapier Nr. 48:

Gundlach, H.: Inventarium der älteren Experimentalapparate im Psychologischen Institut Heidelberg sowie einige historische Bemerkungen (zweite, vermehrte Auflage). (September 1986, Historische Reihe Nr. 9)

Diskussionspapier Nr. 49:

Klüpfel, J. & Graumann, C.F.: Ein Institut entsteht – Zur Geschichte der Institutionalisierung der Psychologie an der Universität Heidelberg. (Oktober 1986, Historische Reihe Nr. 13)

Diskussionspapier Nr. 50:

Drinkmann, A.: Private und öffentliche Self-Consciousness: Eine Zwischenbilanz ihrer empirischen Bewährung. (Oktober 1986)

Diskussionspapier Nr. 51:

Blickle, G. & Groeben, N.: Gegen einen objektivistisch halbierten Kognitivismus: Kognitiv-konstruktives Sprachverstehen und nicht-paradoxe Wirkungen von Lob und Tadel. (November 1986)

Diskussionspapier Nr. 52:

Scheele, B. & Groeben, N.: Eine Dialog-Konsens-Variante der Ziel-Mittel-Argumentation. (Dezember 1986)

Diskussionspapier Nr. 53:

Batz, W.-D., Bickes, C., Bickes, H., Busse, D. & Lörch, B.: Konzeptuelle Strukturen in der Sprache des Vorurteils. (Dezember 1986)

Diskussionspapier Nr. 54:

Röhrle, B.: Soziale Netzwerke und Unterstützung. (Januar 1987)

Diskussionspapier Nr. 55:

Sommer, J.: Der Signifikanztest in der psychologischen Forschung. Ein Falsifikationsinstrument im Sinne des Kritischen Rationalismus? (März 1987)

Diskussionspapier Nr. 56:

Batz, W.-D.: Kodierung und Repräsentation – über hypothetische Mechanismen in Gedächtnistheorien. (Dezember 1987)

Diskussionspapier Nr. 57:

- Bastine, R.: Psychotherapeutische Prozeßanalyse. (September 1987)
- Diskussionspapier Nr. 58:
Amelang, M. & Krüger, C.: Kindesmißhandlung. (November 1989)
- Diskussionspapier Nr. 59:
Amelang, M.: An Investigation of the Factorial Structure and External Validity. (Dezember 1987)
- Diskussionspapier Nr. 60:
Bastine, R.: Klinische Psychodiagnostik. (März 1988)
- Diskussionspapier Nr. 61:
Waller, M.: Die Entwicklung der Beurteilung fehlerhafter Äußerungen – Eine Pilotstudie. (Juni 1988)
- Diskussionspapier Nr. 62:
Schahn, J. & Holzer, E.: Untersuchungen zum individuellen Umweltbewußtsein. (August 1989)
- Diskussionspapier Nr. 63:
Stössel, A. & Scheele, B.: Nomothetikorientierte Zusammenfassung Subjektiver Theorien zu übergreifenden Modalstrukturen. (Januar 1990)
- Diskussionspapier Nr. 64:
Aschenbrenner, K.M., Laier, R. & Albert, D.: Wichtigkeit als Wissen über die Variation der Merkmalsattraktivität bei der Verhaltenswahl. (Dezember 1989)
- Diskussionspapier Nr. 65:
Albert, D., Gertzen, H., Bürgy, R., Bannert, M. & Schneyer, T.: Abruf semantisch strukturierter Informationen beim binären Wählen zwischen beschriebenen Alternativen. (Dezember 1989)
- Diskussionspapier Nr. 66:
Albert, D., Lages, M., Gertzen, H. & Aschenbrenner, K.M.: Beeinflussen Struktureigenschaften von Wissen das Wahlverhalten? (Dezember 1989)
- Diskussionspapier Nr. 67:
Gertzen, H., Bettinger, C., Körner, Chr. & Albert, D.: Bewertende Vergleiche und Informationsabruf in Abhängigkeit von beurteilter Dimensionswichtigkeit bei unvollständig beschriebenen Alternativen. (Dezember 1989)
- Diskussionspapier Nr. 68:
Kane, G., Rotter, B. & Waller, M.: Konstruktion und Erprobung einer Entwicklungsskala zur Erfassung vorsprachlich-gestischer Äußerungen bei geistig behinderten Kindern. Ergebnisse einer Pilotstudie. (Januar 1991)
- Diskussionspapier Nr. 69:
Krüger, C. & Amelang, M.: Arbeitslosigkeit und Kriminalität. (Mai 1991)
- Diskussionspapier Nr. 70:
Groeben, N. & Erb, E.: Reduktiv-implikative versus elaborativ-prospektive Menschenbildannahmen in psychologischen Forschungsprogrammen. (Dezember 1991)
- Diskussionspapier Nr. 71:
Albert, D., Schrepp, M., Held, T.: Construction of Knowledge Spaces for Problem Solving in Chess – Two Experimental Investigations. (März 1992)
- Diskussionspapier Nr. 72:
Kany, W., Waller, M.: Desiderate einer entwicklungspsychologischen Theorie des Spracherwerbs: Eine Positionsbestimmung gegenüber der nativistischen Auffassung Chomskys. (Februar 1992)
- Diskussionspapier Nr. 73:
Kadijk, M.: Plotting Activations in Neural Networks. (Oktober 1992)
- Diskussionspapier Nr. 74:

Unnewehr, J.: Benutzerhandbuch Prozeduren zur Wissensdiagnose. (Dezember 1992)

Diskussionspapier Nr. 75:

Erb, E.: Die Kontraststruktur menschlichen Denkens zwischen Dogmatismus als kurzschlüssiger Polarisierung und polarer Integration als Entwicklungsziel. (Dezember 1992)

Diskussionspapier Nr. 76:

Christmann, U., Dienes, A., Groeben, N., Kolz, A., Schreier, M., Shaery, K.: Ist unser Land wirklich ein 'volles Boot'? (November 1993)

Diskussionspapier Nr. 77:

Erb, E.: Selbstentwicklung/-sozialisation statt Selbstaktualisierung/-verwirklichung? Zur konzeptuellen Kritik und Weiterführung des zentralen Motiv-Konstrukts der Humanistischen Psychologie. (Dezember 1993)

Diskussionspapier Nr. 78:

Schahn, J., Erasmy, P., Trimpin, A., Ditschun, K.: Psychologische Maßnahmen zur Förderung von Hausmüllvermeidung und Hausmülltrennung. (Februar 1994)

Diskussionspapier Nr. 79:

Nüse, R.: New new Representationalism. Eine philosophisch-psychologische Wahrnehmungstheorie. (Juni 1994)

Diskussionspapier Nr. 80:

Schahn, J., Dinger, J., Bohner, G.: Die Rolle von Rationalisierungen und Neutralisationen für die Rechtfertigung umweltschädigenden Verhaltens. (Mai 1994)

Diskussionspapier Nr. 81:

Held, T., Schrepp, M., Fries, S.: Vergleich von Methoden zur Ableitung von Wissensstrukturen. (Dezember 1994)

Diskussionspapier Nr. 82:

Schahn, J.: Umweltpsychologische Bibliographie: Gesamtverzeichnis und nach Themengebieten geordnet. (Januar 1995)

Diskussionspapier Nr. 83:

Korossy, K.: A Qualitative-Structural Approach to the Modeling of Knowledge. (Dezember 1996)

Diskussionspapier Nr. 84:

Schahn, J., Damian, M., Schurig, U., Fuchsle, C.: Konstruktion und Evaluation der dritten Version des Skalensystems zur Erfassung des Umweltbewußtseins (SEU-3). (März 1999)

Diskussionspapier Nr. 85:

Schahn, J. & Bohner, G.: Evaluation Research in Environmental Sociology and in Environmental Psychology: Methodological Aspects. (November 2002)