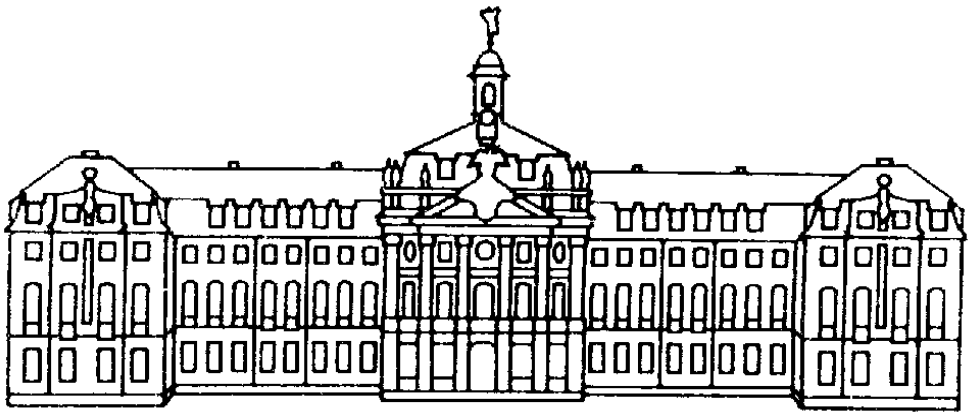


Program evaluation:
A facet-theoretic approach[1]

Anna Döring

2005

Berichte aus dem Psychologischen Institut IV

Sozialpsychologie ♦ Persönlichkeitspsychologie ♦ Organisationspsychologie

Fliednerstr. 21, 48149 Münster

[1] Short version of: Döring, A. (2004*). Program evaluation: A facet-theoretic approach*. Unpublished diploma thesis, University of Münster, Germany.

Die Reihe erscheint von 1986 an in unregelmäßiger Reihenfolge und enthält Forschungsberichte und theoretische Arbeiten von Angehörigen des Psychologischen Instituts IV der WWU, Sozialpsychologie, Persönlichkeitspsychologie, Organisationspsychologie.

Korrespondenzadresse:
Wolfgang Bilsky, Differentielle Psychologie und Persönlichkeitspsychologie, Psychologisches Institut IV der Westfälischen Wilhelms-Universität Münster, Fliednerstr. 21, 48149 Münster, Tel. 0251-83-34198, Fax 0251-83-31343; email: bilsky@psy.uni-muenster.de

Program evaluation:
A facet-theoretic approach

Anna Döring

Contents

1. Introduction

1.1 Program Evaluation

Improving the quality of our physical and social environment and enhancing our individual and collective well-being is a major aim of many programs in various fields of application. But how do we know whether they help us achieve that goal, whether they make sense, whether they are effective?
Activities directed at collecting, analysing, interpreting, and communicating information about the effectiveness of social programs for improving social conditions are called *program evaluation*. Or, in the words of Rossi, Freeman and Lipsey (2002, p.4): “Program evaluation is the use of social research procedures to systematically investigate the effectiveness of social intervention programs.”

Evaluations are conducted for a variety of practical reasons. Once a program has been carried out, an evaluation can aid in decisions on whether the program should be continued, improved or expanded. Or it can help to assess the utility of new programs, initiatives and ideas. In practice, stakeholders want to know whether it is worthwhile investing money and other resources in a specific program. Therefore, the main purpose of program evaluation is to make valid findings about the effectiveness of social programs available to those persons with responsibilities or interests related to their creation, continuation, or improvement.
Research methods from the social sciences are powerful tools for systematically investigating social intervention in its natural political and organizational conditions. They can help us evaluate the design, the implementation, the impact and the efficiency of any kind of measures. In this respect, program evaluation is an application of social research methods to the task of assessing the quality of programs, so that judgements can be drawn for the need of intervention.
Additionally, evaluations may also contribute to substantive methodological social science knowledge, as they help us develop guidelines for good programs. This is exactly the scope of evaluation research: Every single program and every single evaluation can be incorporated into the larger context of other programs in this specific field. Consequently, one can judge the value of this project in comparison to other projects and make use of experiences that you have already made.
So, evaluation research, in the long run, contributes to one of the main aims of science: the accumulation of (scientific) knowledge.

1.1.1 The history of program evaluation

According to Cook and Matt (1990), the crucial point in program evaluation has always been the reduction of complexity. However, how people defined social problems and assessed their impact and which problems were salient to us, changed over time with shifts in values and lifestyles.

Cook and Matt (1990) account for three major periods:

The first period from 1965-1975 is associated with scientists like Donald Campbell and Michael Scriven. To these researchers the discovery of program effects against the background of scientific (experimental) methodology and logic is of great importance. The work of Donald Campbell is carried on by the *Campbell Collaboration* today. For the Campbell Collaboration, field experiments and quasi experimentation, as one of Campbell's contribution to the field, are still the central features of an evaluation study with high methodological quality (Farrington, 2003).

The second period (1975-1982) was characterized by criticism of the merely quantitative and scientific approaches that dominated the first period. Critics doubted the usefulness of a theory of evaluation that was not based on a valid and comprehensive description of how social programs operate and how social scientific knowledge could be used to change programs. Instead, evaluations should satisfy the need for information of all participants. Contributions in this period stress the point of view of practitioners, such as politicians or program managers (see for example Weiss 1973,1975; Wholey 1979, 1983). Although these researchers emphasise one of these points of view or another, they all stress the importance of the social, political and economic context of an evaluation process.

In the third and last period, people tried to combine experiences from the past and to create a theory of evaluation that is less restrictive in methods and the role of the evaluator. A leading person in this period was Peter Rossi. He explained that different parties have an interest in a particular evaluation. An evaluation should take into account this diversity of interests. This approach was called "multi-goal" (Chen & Rossi, 1980).

Aside from the diversity of interests, he preferred a diversity of research styles. For Rossi et al. (2002), a good evaluation does not necessarily use one specific design (such as Campbell's quasi experiment), but the best design for that specific purpose.

Rossi (Chen & Rossi, 1983) stated that the domination of the experimental paradigm in program evaluation literature has unfortunately drawn attention away from a more important task of understanding social programs, namely, developing theoretical models of social intervention.

1.1.2 Program Evaluation today – important sources and contributions

Today, there is a vast amount of literature on program evaluation. However, some works are of outstanding importance in current research.

Rossi's book *Program Evaluation – A systematic approach* (Rossi et al., 2002) is certainly one of the most important evaluation guidelines today. The book gives an introduction to activities used in judging the design, the implementation, and the utility of social programs. It communicates the technical knowledge and collective experiences of practicing evaluators. Its approach is a multidisciplinary one, and it addresses various audiences, practitioners and sponsors of social programs, as well as scientists.

An important contribution to today's knowledge on program evaluation is made by another approach: During the last decades, much research has been done on the development of internationally accepted standards of program evaluation. In 1975, a committee appointed by the American Educational Research Association, The American Psychological Association, and The National Council on Measurement in Education found that there was no clear definition of what constitutes a reasonable evaluation of educational programs. Therefore, it compiled knowledge about program evaluation gained from professional literature and from years of experience by educators and evaluation specialists. This knowledge was organised and presented in a book on standards for the practice of educational program evaluation (Joint Committee on Standards for Educational Evaluation, 1994).
The Joint Committee (1994, p.3) defines an evaluation standard as follows: "A principle mutually agreed to by people engaged in the professional practice of evaluation, that, if met, will enhance the quality and fairness of an evaluation."
The *Standards* are a comprehensive framework that should be helpful in a vast amount of situations. The Joint Committee holds the view that good evaluations require the involvement of many people with different perspectives.
Since then, the *Standards* have been continuously improved. They have been translated into various languages. The vast amount of references to the standards in journal articles in central evaluation journals and on international conferences emphasises their importance.

1.2 Program evaluation in the field of crime prevention

Recently, effort has been made to develop evaluation standards for one specific content area: the field of crime prevention. Important questions in this context are: What works to reduce crime? How should offenders be dealt with so that they do not reoffend? What methods of preventing crime are most cost-effective?
Organizations, often supported by governments or municipalities, have developed guidelines especially for the everyday practice of evaluation (e.g., Home Office, 2002; Zentrale Geschäftsstelle Polizeiliche Kriminalprävention der Länder und des Bundes, 2003). Apart from this, there are hundreds of crime prevention programs that have already been evaluated. An important concern of governments and municipalities was to take advantage of this huge amount of existing evaluation studies. On the basis of existing evaluation studies, they intended to determine what works in order to reduce crime. Therefore, the results of many evaluation studies had to be reviewed. Recently, two reviews of crime prevention projects have been conducted.

1. In 1996, the Congress of the United States of America required the Attorney General to provide a “comprehensive evaluation of the effectiveness” of over $3 Billion annually in Department of Justice grants to assist State and local law enforcement and communities in preventing crime. Congress required that the research for evaluation “employ rigorous and scientifically recognized standards and methodologies.” In comparing evaluation studies from various content areas and applying different approaches, this report directly confronted the problem of mixed results. This task was solved in the so called “Sherman Report” (Sherman et al., 1997).
2. An equivalent to the Sherman Report in the United States is the so called “Düsseldorfer Gutachten” in Germany (Landeshauptstadt Düsseldorf, 2001). Here, several hundred evaluation reports from different countries are compared with each other.

In these two reviews, evaluation research focuses on a summary of the evaluation of diverse programs in a descriptive text. Especially in the Düsseldorfer Gutachten, the review is conducted within a relatively idiosyncratic format. Each program is described in an essayistic, non-formalised style. As every program is described separately, a comparison between the programs remains difficult. This is why this type of review faces the problem that readers often cannot determine how and why reviewers drew their conclusions about what works.

1.3 The usual approaches: systematisation and integration

In order to overcome this kind of problems and to offer enhanced formalisation, two techniques have been developed. While the first technique aims at enhanced systematisation, the second technique aims at summarizing research findings.
Two international organisations, the Cochrane Collaboration for the field of health care interventions and the Campbell Collaboration for the field of crime prevention measures, were concerned with the systematisation of reviews. They have developed the “science of systematic reviews” (Farrington, 2001, p.127). Systematic reviews have explicit objectives and explicit criteria for including or excluding studies. They use rigorous methods for locating, appraising and synthesizing evidence from prior evaluation studies.
The claim of the Campbell Collaboration is to include only studies with high methodological quality in a review. But what are the features of an evaluation study with high methodological quality? In trying to specify these features for criminology and the social and behavioural sciences, the Campbell Collaboration focuses on the work of Donald Campbell and his colleagues (Campbell and Stanley 1966; Cook and Campbell 1979; Shadish, Cook, and Campbell 2002). Campbell was one of the leaders of the tradition of field experiments and quasi experimentation (Shadish, Cook, and Campbell 2002). Accordingly, the main focus is on validity and methodological accuracy.
The second contribution to improving reviews was made by a technique called “meta-analysis” (Lipsey & Wilson, 2001; Rosenthal, 1983). In a meta-analysis, empirical findings of multiple studies that research the same issue are synthesised. The purpose is to develop generalisations about research findings across different studies, to look for common trends and findings. This is accomplished by translating the results of different studies into effect sizes, which represent the effect magnitude of the results. Meta-analysis offers procedures to calculate mean effect sizes for one crime prevention measure across different studies, so that reliable effects can be calculated beyond the relatively small sample in a single evaluation study. This is one way to provide a common framework for existing research that allows for comparisons and generalisations.

1.4 Facet Theory: An alternative approach

Another approach that takes “the stance of making the most effective sense possible of the existing publications and data sets available” is facet theory (Canter, 1985, p. ix). Facet theory (FT), an approach developed by Louis Guttman in 1959, provides scientists with a language to talk about their domains of discourse, and to structure and analyse the

instruments they use to study this domain. It is concerned with the integration of concepts and data to facilitate the discovery of lawfulness in complex systems. For Guttman (1959) a methodological approach that is concerned with relations between concepts is called a *theory*. Facet theory, as a theory about research activities, is a theory about how theories themselves can be specified and tested (Canter, 1985). Facet theorists relate different concepts to form a theoretical framework. This framework helps them to look for trends that are common across a number of separate research activities and to demonstrate that certain patterns can be consistently identified. Consequently, everybody who uses facet theory can be called a meta-analyst in some sense.
Using FT means formalizing knowledge, finding a structure in a certain content area and summarising the main ideas. Formalisation of research is helpful for communicating knowledge within the scientific community for objective evaluation. Additionally, it helps stating ideas with minimal ambiguity and minimal redundancy, and to focus on the central aspects under study. FT methods are geared toward generalisability and cumulative theory construction; they aim at explicating the implicit. In this way, it becomes clear that evaluation research and FT share many of their main objectives.

The basic idea of this diploma thesis is to get away from the still predominant, more or less essayistic, type of evaluation reviews by making the criteria applied more obvious and comprehensible. Facet theory, that in fact has never been applied to the field of crime prevention, seems to be a powerful tool in trying to solve the problems associated with traditional evaluation reviews. Facet theory takes into account the systematisation of evaluation reports as proposed by the Campbell Collaboration as well as the integration of different studies as one of the main aims of meta-analysis.

1.4.1 Important concepts in facet theory

In this paragraph, I will give a brief introduction to the main concepts in facet theory. For details, one might refer to Canter (1985) or to Shye and Elizur (1994) who provide a comprehensible and comprehensive description.

FT provides practical tools for all phases of scientific inquiry: formulating hypotheses, analysing and interpreting data and presenting results. In FT, models are combined with data analytic techniques and visual displays of the structure of data.

Facets

The building blocks of facet theory are *facets*. Facets offer a way to systematise scientific research: Typically, scientific research starts with a collection of variables that might be important to describe objects of interest. Guttman suggests that our perception or thinking about this collection of variables should lead to identification of semantic or perceptual properties that characterise basic components of relevant variables (Dancer, 1990). These properties are formulated in facets, sets of elements that classify objects under study. Facets make distinctions that are, by definition or by hypothesis, relevant for the scientific investigation (Borg & Shye, 1995).

The facet-theoretic model: mapping sentences

To aid in communicating, facets are represented in the context of a sentence, called a *mapping sentence*. A mapping sentence specifies the relations between the facets in a sentence which is formulated in everyday language. In this way, it puts each facet into a broader context. In other words, the formulation of the facets and the specification of the relations between facets is the facet-theoretic approach towards structuring a certain content area.
As in any other research framework, a facet theorist is interested in certain samples of persons or objects. The special feature of the facet approach is its ability to describe each member of the sample in terms of the facets and their interrelations.

The mapping sentence as the facet-theoretic model plays a crucial role in a facet-theoretic investigation. The facet-theoretic point of view on the proceeding in scientific research was formulated by Louis Guttman (Levy, 1994). For him, an empirical theory is a "*hypothesis of a correspondence between a definitional system for a universe of observations and an aspect of the empirical structure of those observations, together with a rationale for such an hypothesis"* (p.63). This is summarised in a mapping sentence:
Based on theoretical knowledge in a specific content area, the facet theorist identifies facets and specifies their interrelations in a mapping sentence.
Furthermore, the mapping sentence shows the relationships between the facets and the scale on which values are attached to members of the sample (Borg, 1996). In this way, assumptions about structural relations are linked to concrete data. On the basis of the facets and their interrelations, the facet theorist formulates hypotheses about the configuration of the data.

Finally, the structure within the data can be compared with his or her theoretical expectations that have been formulated in the mapping sentence. Accordingly, one of the central questions in a facet-theoretic study is: Can the structure of the data be explained by the facets?

1.4.2 Types of data

In evaluation research, the vast amount of information is conveyed by words rather than by numbers. Except for calculations of effects or the use of statistics for descriptive purposes, the evaluation of programs deals only with descriptions, explanations, arguments or any other type of investigation with the help of words.
Facet theory can deal with these data. Canter (1985) describes two advantages of using facet theory for the analysis of qualitative information:
First, by developing faceted definitions it is possible to provide a formal and reasonably precise mechanism for specifying the content categories around which the qualitative material may be examined.
The second way in which the facet approach helps with qualitative information is by providing analysis procedures which do not require high levels of measurement in order to function effectively.

1.4.3 Data analytic techniques

Facet theory emphasizes the role of mathematics as a formal language, efficient for communicating certain aspects of reality. With sharply defined facets and an enhanced technicality of the facet's interconnections, mapping sentences become, in effect, mathematical models. It offers a series of inventive data-analytic techniques. These data-analytic techniques can show fruitful representations in geometric form (Canter, 1985). Spatial patterns are integrated into an elaborated theoretical framework. In this way, FT is really a methodological approach (Borg & Shye, 1995), and the permanent interaction between theory and results eases the task of theory development.

1.5 Rationale of this thesis

In this thesis, I am interested in taking advantage of the general potential of facet theory by applying it to the field of evaluation, i.e. facet theory is applied to the evaluation of crime prevention measures as one specific content area that is in the focus of current research and political discussions.
As can be derived from the comparison of facet theory with the "science of systematic reviews" (Campbell) and meta-analysis (see above), facet theory seems to be a powerful approach towards the systematisation of evaluation as well as towards the integration of evaluation studies. To systematise evaluation and to integrate evaluation studies, the present research is based on previous findings, namely on a facet-theoretic model for evaluation that has recently been developed.

1.5.1 A facet-theoretic model for evaluation

The first attempt to formulate program evaluation standards in terms of a facet-theoretic model was made by Bilsky and Cairns (in press). They have developed the following mapping sentence for program evaluation, which forms the starting point for my research:

X: person			*A: intervention*		*B: time*	
(x1 practitioner)					(b1 finished)	
(x2 evaluator)	evaluates a		(a1 project)	which is	(b2 in progress)	
(x3 stakeholder)			(a2 program)		(b3 planned)	
(x4 other)					(b4 unspecified)	
	C: standard		*R evaluation*			
	(c1 utility)		(very positive)			
with respect to	(c2 feasibility)	→	(…)	evaluation in the sense of facet C.		
	(c3 propriety)		(very negative)			
	(c4 accuracy)					

This model contains the typical elements of a mapping sentence:

- A facet for the person (x)
- Several domain facets (A, B and C) that are specific for a certain content area (i. e. intervention, time, standard)
- A response facet (i.e. evaluation) that specifies the range of possible values.

Each facet describes one important aspect of an evaluation. The role a certain facet plays is clarified by its placement in the mapping sentence and the corresponding relation to the other facets. In terms of the mapping sentence, each evaluation can be described by a

combination of elements, one element from each facet. The unique combination of elements that characterise an evaluation has practical implications.

Firstly, it is important for the design and the result of an evaluation which perspective it is conducted from, as choice of criteria and their emphasis can differ dependent on the respective evaluator (facet x: person).

More, there is always an object of evaluation, that is one unique intervention under study (facet A). Against the background of limited resources (time money, personnel), it is important to know whether the temporal extent of a measure is limited (project) or whether it is continuous (program). This facet is expected to moderate the outcome of an evaluation. However, the present study will focus on aspects of the evaluation process itself, so that the distinction between project and program is of minor importance. The two expressions will be used interchangeably throughout the thesis.

Finally, the time relation as a central framework of an intervention is explicated in facet B (time).

The facets that have been mentioned so far supply a description of an evaluation by referring to external factors which do not provide the researcher with direct access to the proceeding of evaluating. However, what researchers are most interested in (e.g., Rossi et al., 2002; Joint Committee, 1994) is the way in which an evaluation is conducted. Standards or guidelines to judge the quality of an evaluation are in the focus of current research. For this reason, facet C (standard) is the core of the mapping sentence. Its importance is visualised as it is directly referred to in the response facet.

The notation of the elements in this facet corresponds to the notation within the *Standards* developed by the Joint Committee (Joint Committee, 1994). The standards are organized around the four important attributes of an evaluation: *utility, feasibility, propriety, and accuracy*.

The Joint Committee holds the view that these four attributes are necessary and sufficient for a sound and fair evaluation.

Utility	Utility standards guide evaluation so that they will be informative, timely, and influential. They require evaluators to acquaint themselves with their audience, define the audience clearly, ascertain the audiences' information needs, plan evaluations to respond to these needs, and report the relevant information clearly and in a timely fashion. (p.5)
Feasibility	Feasibility standards recognize that evaluations usually are conducted in a natural, as opposed to a laboratory, setting and consume valuable resources. Therefore evaluation designs must operate in field settings, and evaluations must not consume more resources, materials, personnel, or time than necessary to address the evaluation questions. (p.6)

Propriety	Propriety standards reflect the fact that evaluations affect many people in a variety of ways. These standards are intended to facilitate protection of the right of individuals affected by an evaluation. They promote sensitivity to and warn against unlawful, unscrupulous, unethical, and inept actions by those who conduct evaluations. (p.6)
Accuracy	Accuracy standards determine whether an evaluation has produced sound information. The evaluation of a program must be comprehensive; that is, the evaluators should have considered as many of the program's identifiable features as practical and should have gathered data on those particular features judged important for assessing the program's worth or merit. Moreover, the information must be technically adequate, and the judgements rendered must be linked logically to the data. (p.6)

Especially the accuracy standard is often in the focus of evaluation literature (e.g., Rossi, Freeman & Lipsey, 2002).

1.5.2 Hypotheses

This facet-theoretic model of Bilsky and Cairns is the first and only mapping sentence dealing with program evaluation standards so far. However, it has not been tested empirically yet. Furthermore, facet theory has never been applied to crime prevention, so that little is known about the area under investigation in this study. This is why the present research starts with very general expectations:

First, I am interested in systematising evaluation research by using facet theory, i.e., in formulating evaluation research in terms of facet theory. Consequently, the first goal is to develop and to test a facet-theoretic model for evaluation.

Second, I am interested in integrating concrete evaluation studies into a common facet-theoretic framework. In this way, one of the main aims in current research on evaluation (see for example Rossi et al., 2002; Joint Committee, 1994) should be achieved: the assessment of the quality of evaluation studies.

The general expectations for this thesis are formulated in the following hypotheses:

Hypothesis 1:	It is possible to state evaluation research actions in terms of facet theory to reach conceptual clarity.
Hypothesis 2:	Within a facet-theoretic framework, the quality of evaluation reports can be assessed, so that a distinction can be made between well-done and not well-done evaluations.

1.5.3 Approach

This thesis centres around a test of these hypotheses.

Investigation of hypothesis 1

As described above, two distinct aspects are of importance to the investigation of the first hypothesis. At first, evaluation research actions are stated in terms of facet theory. Based on this theoretical part, it will be investigated whether a statement in terms of facet theory leads to conceptual clarity, which will be investigated empirically. This is why the first hypothesis will be investigated in two steps. While the first step deals with the development of a facet-theoretic model, the second steps deals with the empirical test of the model.

Step 1: Development of a facet-theoretic model

The mapping sentence of Bilsky and Cairns (in press) will be specified step by step. Each mapping sentence on the next lower level will become more concrete, so that a set of hierarchically structured mapping sentences will evolve. The mapping sentences will be developed on the basis of central program evaluation literature.
The development follows Rossi's (Rossi, Freeman & Lipsey, 2002, p.22) "systematic approach" towards program evaluation. This means that the model should evaluate an evaluation report according to its agreement with the rules of social research procedures, nevertheless allowing for flexibility in methods.

Step 2: Test of the model

For an empirical test of the model, an evaluation report of crime prevention measures is chosen as the unit of analysis. An evaluation report, as the final product of every sound evaluation of a program, has been introduced as the standard unit of analysis in evaluation research (see above). It usually contains information about the program, the proceeding and the results.
Evaluation standards form the core of the initial mapping sentence of Bilsky and Cairns. Following this emphasis, standards for evaluation will play the central role in the test of the model in the present study. Two aspects are of importance: suitability of the standards for analysis and investigation of the structure of these standards.

The evaluation standards formulated in the model should be suitable for further analysis. This is to say that the application of these standards makes sense in so far as it provides the researcher with information. This will be clarified by using descriptive analyses.
The suitability of the standards for analysis is a prerequisite for the investigation of the central question concerning the structure, i.e.: What are the relationships between evaluation standards? This will be investigated by taking advantage of facet-theoretic methods of analysis.

Investigation of hypothesis 2

At this stage, it is investigated whether a facet-theoretic model and facet-theoretic data analyses are useful tools for the researcher who is concerned with the assessment of the quality of evaluation reports. The central aspect under study is the usefulness of evaluation criteria in assessing the quality of reports. This means, the analyses should show whether it is possible to distinguish qualitatively good evaluation reports from qualitatively bad ones by applying the chosen criteria.

2. Method

The investigation in the present study centres on a test of the two hypotheses that are formulated in the Introduction. As explained in the Introduction, the test of the first hypothesis stresses two distinct aspects. First, a facet-theoretic model for evaluation is *developed*. In the next step, this model is *tested empirically*. This empirical test accentuates *applicability* on the one hand, and *structure* on the other hand. The test of the second hypothesis aims at the *assessment of the quality of evaluation reports* within a facet-theoretic framework.

This part of the thesis is structured in accordance with these three aspects, namely, the development of a facet-theoretic model, the empirical test of the model, and the assessment of the quality of evaluation reports within a facet-theoretic framework.

2.1 Design

2.1.1 Sample

A general mapping sentence for evaluation could be applied to any evaluation report. For illustrative purposes, a specific sample of programs is drawn. The sample of this study consists of evaluation reports which describe the implementation and evaluation of crime prevention programs.
The approach that is applied here is systematic sampling; that means I systematically choose crime prevention programs that have been evaluated in a more or less advanced way. The rationale for this proceeding is as follows:
As the aim of this study is to develop and apply a model that distinguishes good evaluations from poor ones, it is important to cover a broad band width of quality. Furthermore, I have to make sure, that the reports contain enough information about the evaluation itself to make an application of the model fruitful. A review of a report that has been conducted by a methodological expert ensures a minimum level of quality. Additionally, a report can only be reviewed if it contains enough information. For these reasons, the following eligibility criterion is established:
Only those reports are chosen that are included in the “Düsseldorfer Gutachten” (2000) or in a review conducted recently by Martin Süß from the Otto-von Guericke University of Magdeburg (Süß, 2003).

Beside this, the emphasis is put on increasing variation, in order to ensure that the variability is sufficient, to reveal the basic lawfulness sought. Therefore, the sample is drawn so that a high level of heterogeneity is reached.

2.1.2 Development of a facet-theoretic model – a set of hierarchically structured mapping sentences

In designing a model, there are basically two aspects to consider: the content and the form of the model.

Content

A facet-theoretic model of program evaluation should be based on up-to-date knowledge about program evaluation. Additionally, sources for the content of the model should be internationally accepted and based on scientific expertise. As explained in the Introduction (pp. 3-4), there exist internationally accepted program evaluation standards of outstanding importance: *The Program Evaluation Standards* by *The Joint Committee on Standards for Educational Evaluation* (1994). These standards form the basic framework of my model.

The four main attributes in the Standards of the *Joint Committee* (1994), *Utility, Feasibility, Propriety and Accuracy*, are a central part of the mapping sentence of Bilsky and Cairns. However, most of the information that is given the evaluation reports in the Düsseldorfer Gutachten and in the review from Süß can be assigned to the attribute *Accuracy*. Accuracy is over-represented whereas *Utility*, *Feasibility*, and *Propriety* seem to play a minor role in reporting program implementation and evaluation. This is not to say that these features are not important in the proceeding of an evaluation. But, as the source of data in this study is a sample of evaluation *reports*, it seems to be most promising to focus on accuracy standards.

The section on Accuracy in the *Standards for Program Evaluation* covers 12 standards:

1. **Program Documentation:** The program being evaluated should be described and documented clearly and accurately, so that the program is clearly identified.
2. **Context Analysis:** The context in which the program exists should be examined in enough detail, so that its likely influences on the program can be identified.
3. **Described Purposes and Procedures:** The purposes and procedures of the evaluation should be monitored and described in enough detail, so that they can be identified and assessed.
4. **Defensible Information Sources:** The sources of information used in a program evaluation should be described in enough detail, so that the adequacy of the information can be assessed.
5. **Valid Information:** The information gathering procedures should be chosen or developed and then implemented so that they will assure that the interpretation arrived at is valid for the intended use.
6. **Reliable Information:** The information gathering procedures should be chosen or developed and then implemented so that they will assure that the information obtained is sufficiently reliable for the intended use.
7. **Systematic Information:** The information collected, processed, and reported in an evaluation should be systematically reviewed and any errors found should be corrected.
8. **Analysis of Quantitative Information:** Quantitative information in an evaluation should be appropriately and systematically analysed so that evaluation questions are efficiently answered.
9. **Analysis of Qualitative Information:** Qualitative information in an evaluation should be appropriately and systematically analysed so that evaluation questions are effectively answered.
10. **Justified Conclusions:** The conclusions reached in an evaluation should be explicitly justified, so that stakeholders can assess them.
11. **Impartial Reporting**: Reporting procedures should guard against distortion caused by personal feelings and biases of any party to the evaluation, so that evaluation reports fairly reflect the evaluation findings.
12. **Metaevaluation:** The evaluation itself should be formatively and summatively evaluated against these and other pertinent standards, so that its conduct is appropriately guided and, on completion, stakeholders can closely examine its strength and weakness. (pp. 125-126)

As an evaluation report deals with the evaluation of a program and not with meta-evaluation, the twelfth standard will not be included in the model.

The *Standards* are presented in a uniform format to facilitate reading and to illuminate their interrelationships. Each standard is described by a descriptive title and an overview text. The overview text is a conceptual statement that gives definitions of key terms and a general rationale for its use.
For the purpose of collecting variables for the model, the overview texts for each of the 11 standards are scanned for key terms. All key terms will be included as variables in the model.

However, the Standards do not supply the evaluator with detailed technical standards. They are not intended for replacing textbooks in technical areas such as qualitative and quantitative research design and analysis, measurement and data collection, data processing, and report writing (cf. Joint Committee, p. 2)

For the purposes of the present model, this means that the first standard *Program Documentation* is covered by a very broad overview section. Besides, when addressing the validity of inferences in an evaluation, the overview section provides general guidelines for determining the validity of inferences. However, it does not present detailed validation procedures. For these reasons, additional background information is necessary.

For the standard *Program Documentation*, I choose Rossi et al. (2002) to complement the key terms taken from the overview text in the *Standards for Program Evaluation*. Basically, the model refers to two criteria: program theory and goals/objectives.
"Program theory" (p.155) is the conception of the program. It is "based on a thorough understanding of the social problem the program is intended to address" (p.174).
"The programs goals and objectives (should) be specified" (p.174), and there should "be observable implications of the goals and objectives such that meaningful measures and indicators of success could be defined". (p.179)

For the standard *Valid Information* additional information about types of validity is necessary. The Campbell Collaboration emphasises the role of validity and provides practical guidelines for the assessment of validity in an evaluation report. Current guiding principles for the evaluation of crime prevention measures are explained in a journal article published by the Campbell Collaboration (Farrington, 2003), which seems to be especially suitable, therefore. According to Cook and Campbell (1979) and Shadish, Cook and Campbell (2002), methodological quality depends on four criteria: *statistical conclusion validity*, *internal validity*, *construct validity*, and *external validity*. This validity typology has always been the central hallmark of Campbell's work over the years (Shadish, Cook, and Campbell 2002).
External validity refers to the generalisability of causal relationships across different persons, places, times, and operational definitions of interventions and outcomes (e.g., from a demonstration project to the routine large scale application of an intervention). It is difficult to assess this within one evaluation study, unless it is a large-scale, multisite trial. External validity can be established more convincingly in systematic reviews and meta-analyses of numerous evaluation studies. Therefore, external validity is not included in the model.

Form

Apart from compiling a set of variables of outstanding importance, the design of a good model aims at a clear presentation of these variables. The principle question is: How can the

set of variables be represented most elegantly in a mapping sentence? How are these variables best presented in one or several facets?

To answer this question according to facet-theoretic proceeding, one has to consider the intrinsic nature of the structure of variables. What do the variables have in common? Do they overlap content-wise?

All variables describe different aspects of good evaluation. A higher value on each of the variables represents a better evaluation than a lower value. Each report can be classified on each of the variables. The result of this classification will be a profile for each report, consisting of values on each variable.

In facet-theoretic terms, this can be stated as follows: A set of variables from which profiles are generated measures a single common construct. The variables in this study should be chosen so that they do not overlap with respect to their content. In this way, the model becomes economical.

The presentation of the variables in the mapping sentence of the present study follows models from other content areas where the structure of variables is the same.

Shye (1985, p.63, p.64, p.87, p.155, p.219) presents several mapping sentences with sub-sets of variables that have the following features, and, hence, are similar to the present set of variables in their intrinsic structure: Firstly, objects (in Shye's text: persons; in the case of my study: evaluation reports) are categorised by a profile on a set of variables. Secondly, these variables do not overlap in their content. And finally, the variables all measure a single common construct (in the present study: qualitatively good evaluation).

Levy recently presented a problem which is structured similarly (Levy & Bar-On, 2003). The mapping sentence that she proposes for this problem (see Appendix A) follows the same structure as the mapping sentences in Shye (1985).

All these mapping sentences present the subset of variables in one single facet. Following these ideas, the set of variables for evaluation criteria in the mapping sentence in this study is presented in one facet, too. This facet contains variables that do not overlap in their content. Nevertheless, it has to be mentioned that this facet does not meet an expectation that is often emphasised in facet theory: A facet should be a set of mutually exclusive categories (Canter, 1985). Borg (1996), however, explicates that the appropriate level of formalisation depends on the knowledge in this specific area. And this is the first attempt to elaborate a model. Maybe, this facet will be formulated more elegantly with increasing knowledge about the structure of the variables within this facet.

2.1.3 Empirical test of the hypotheses – preparatory work

Development of an instruments for data collection

In order to enable the researcher to test the two hypotheses empirically, the model has to be applied to concrete evaluation reports. Still, the model itself is stated in relatively abstract terms. That means, it does not supply rules for the measurement of variables. Therefore, the researcher uses an instrument that serves as link between the model and the process of data collection: a *coding frame*.
The purpose of the coding frame is to make the application of the model to the reports more concrete and to give explicit guidelines. In this way, the application of the model should become more reliable, and it should be avoided to give too much leeway to the coding person.
The principle of the coding frame is to allow for measurement of each of the variables within the facet for evaluation standards, the central facet of the model. This is accomplished by formulating *indicators* for each variable, which express the basic ideas of the variables in concrete terms. Basically, the formulation of the indicators follows the same literature as the mapping sentence. However, this literature is just intended for providing a general framework for designing and assessing evaluations, not for replacing textbooks, manuals and handbooks concerned with evaluation instruments and methods. For this reason, the formulation of the indicators is based on additional sources:

- An internationally accepted, up-to-date book about statistical methods: Howell (2002)
- The *SPSS* Output as a standard for displaying the results of statistical tests.

To facilitate the coding process, the coding frame (Appendix B) is complemented by a coding sheet, which highlights the rules for coding formulated in the coding frame. Additionally, it offers space to fill in concrete values for coding a report. The coding sheet is part of Appendix B.

Intended creation of data matrices

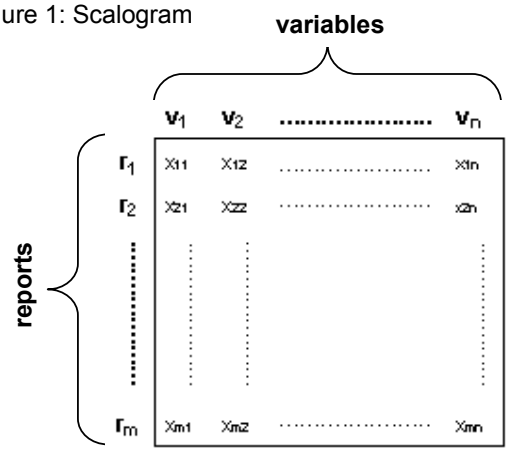
The intended measurement of the variables by using the coding frame can be described as follows:
Each evaluation report is coded on each of the indicators. On each indicator, a report can get a value of either zero or one, with one indicating a better quality of the report. The number of

indicators per variable directly follows suggestions in the literature. Accordingly, there can be more than one indicator per variable.
Then, the values on the respective indicators are summed up, so that the value of a report on a certain variable corresponds to the sum of the values on the respective indicators. Accordingly, the range of possible values per variable varies according to the number of indicators. The level of scaling for these variables is *ordinal*, a higher value indicates a better quality.

Apart from these ordinal variables, the present research is based on an additional type of data: dichotomous data. For this purpose, the levels of the ordinal variables are classified as belonging to one of two distinct categories. In this way, the ordinal variables are dichotomised, so that a report can receive a value of either zero or one on each variable, with one indicating a better quality.

The two data sets are presented as a set of profiles. There is an individual profile for each of the reports, consisting of one value per variable. Each evaluation report can be described as a combination of values on the variables. This is called “structuple” in terms of facet theory. These profiles can be summarised in a rectangular matrix, called “scalogram”, in which the evaluation reports are presented in the rows, and the variables are presented in the columns (Figure 1).

Figure 1: Scalogram



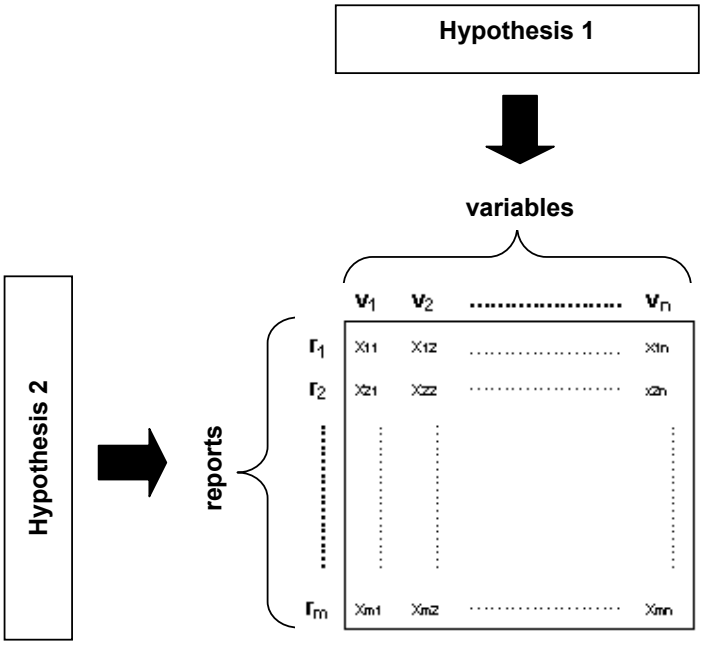
Consequently, two matrices are used as data sets in this study: One matrix with the initial ordinal variables, and a second matrix with the dichotomised variables. The two matrices and a detailed explanation of the process of dichotomisation are presented in Appendix C. The choice of the respective data set for the investigation of the two hypotheses, will be based

on the rationale of the procedure and on the prerequisites of the respective data-analytic method.

2.1.4 Empirical test of the hypotheses: Different perspectives on the data matrix

The empirical investigation of both the first and the second hypothesis is based on the same data matrix described above (Figure 2). It is the perspective that makes the difference. While the test of the first hypothesis focuses on the investigation of the variables, the test of the second hypothesis focuses on the structure of the reports.

Figure 2: Two different perspectives on the data matrix



2.1.5 Empirical test of hypothesis 1

Suitability of the evaluation criteria for analyses

The present research aims at developing a model that is applicable to evaluation reports. Here, the analyses go beyond the mere applicability of the coding frame. The central question is: Does the application of the evaluation criteria under study to the chosen reports make sense? A central quality feature of suitable evaluation criteria is their usefulness in giving information on differences between reports. Only those evaluation criteria that provide

the researcher with information are promising for further analyses. This is investigated by using descriptive analyses. The descriptive analyses focus on the distribution of reports in the chosen variables. To ensure that the variables can distinguish between the reports, a certain amount of cases in each category of a variable is necessary. The following eligibility criterion was established: After dichotomising, a variable should have at least a relation of 20% to 80% in its categories. That means, having less than 5 reports in one category or having more than 20 reports in one category respectively is not acceptable. Variables that are not useful in providing the researcher with information about the reports, are not promising for further analyses. Therefore, variables that do not meet these expectations are excluded and not used for further analyses.

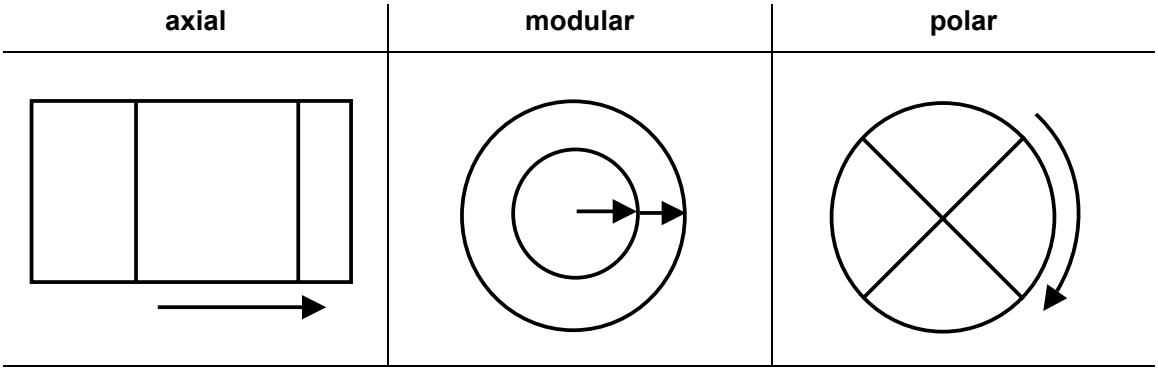
Analysis of the structure – Revealing the relationships between evaluation criteria

The analysis of the structure of the variables concentrates on revealing the relationships between evaluation criteria. This is accomplished by applying a *Smallest Space Analysis* (SSA) to the two data sets. Smallest Space Analysis or Similarity Structure Analysis (SSA), also referred to as Multidimensional Scaling (MDS), is a technique for the analysis of similarity or dissimilarity data on a set of variables. SSA attempts to model measures of proximity for these variables as distances among points in a geometric space. The main reason for doing this is that one wants a graphical display of the data structure, one that is much easier to understand than an array of numbers and, moreover, one that displays the essential information in the data, smoothing out noise (Borg & Groenen, 1997). The statistical background is based on geometrically portraying intercorrelations among variables in a Euclidian space. Each variable is presented as a point. The distances between variables correspond to their intercorrelations. Hence, two points are closer if the correlations between the corresponding variables are higher (Guttman, 1968).

SSA provides a tool for the evaluation of the goodness of a solution: a coefficient of alienation as an indicator of the fit between the intercorrelations of variables and their display in the space. Generally, any coefficient between 0 and 1can be assigned to an MDS solution, where 0 stands for a perfect solution. The lower the coefficient, the better the solution. Louis Guttman required that the coefficient of alienation *K* should be less that 0.15 for an acceptable precise MDS solution. However, if the number of variables clearly exceeds the number of dimensions, higher coefficients will also be acceptable (Borg & Groenen, 1997). The evaluation of the quality of the MDS solution in the present study will be based on these clues.

MDS can be used as an exploratory technique to represent the interrelations among variables. Another possibility is to test structural hypotheses. In this study, a combination of both approaches is applied.
In an MDS, variables are arranged in a space, which can be partitioned by a facet in various ways. Most frequently, facets play *axial*, *modular* or *polar* roles (Borg, 1992; Borg, 1996; Dancer, 1990; Levy, 1985). While a facet that plays an axial role partitions the space into parallel slices, a modular facet is represented by circular bands around a common origin. If a facet plays a polar role, it will partition the space into wedge-like regions emanating from a common origin (Figure 4).

Figure 4: Partitioning of the MDS-space by facets



The main distinction among these roles is related to the property of order among elements of the facet. Facets whose elements represent ordered, quantitative attributes of a content universe are said to be ordered facets, and each successive element in the ordering indicates a greater amount of the attribute than the preceding element. Ordered facets can play an either axial or modular role (Dancer, 1990).
A facet whose elements represent unordered qualitative aspects of a content universe plays a polar role and partitions an MDS space into wedge-shaped regions (Dancer, 1990). The variables in the present study deal with different aspects of the evaluation process, so that the content should not be ordered. Some variables are closer together according to their content, some are far away from each other. Regions found to be adjacent to one another will correspond to elements of facets that are more similar conceptually than are elements that correspond to non-adjacent regions. Consequently, it is hypothesised that the variables will be arranged in a circular order and that the space can be partitioned into wedge-shaped regions (see Figure 4). However, the precise circular order of the elements is not known and should be explored with the help of the computer program *SYSTAT 10*. Both data sets are used in this analysis, the data set with the ordinally scaled variables as well as the data set with the dichotomised variables. According to the scale level of the variables under

investigation, measures of proximity are chosen, which serve as a basis for displaying the variables in a geometric space. The coefficient *MU2* was chosen for the ordinal data, the *Jaccard* coefficient for the dichotomous data. The results for both data sets will be compared with each other.

2.1.6 Empirical test of hypothesis 2

Distinction between good and poor reports

Investigating hypothesis 2, it should be tested whether the quality of reports can be assessed by the criteria formulated in the facet-theoretic model. In this part of the study, it is not the structure of the variables that is in the centre of interest, but the structure of the reports. The purpose is to create displays of the structure of the reports that visualise their quality. Referring to these displays, the usefulness of each variable in distinguishing between qualitatively good and qualitatively bad reports will be investigated.
In facet theory, this can be accomplished by using *Multiple Scalogram Analysis* (MSA). The distinction between MSA compared with SSA is that whereas SSA aims at establishing *structural properties of variables*, MSA investigates *structural characteristics of subjects* (Guttman, R. & Greenbaum, 1998), or in this case, reports.
The algorithm of MSA attempts to represent the reports as points in a geometric space such that the space can be partitioned by each and every variable (cf. Borg & Shye, 1995). This is exactly the purpose of the investigation of the second hypothesis. Besides, the interpretation of the results of an MSA is very comprehensible. Hence, it seems to be an ideal method in the field of evaluation, where comprehensibility is very important with regard to an effective communication between scientists and practitioners. For these reasons, MSA is chosen as the basic method.

MSA creates a representation of the reports as points in a space, so that there will be a partitioning of the space into regions according to the values of each variable (Wilson, 1995). Each report appears as a *point* in the space, each variable as a *partition* of the space into non-overlapping regions, and each category as a *region* (see Zvulun, 1978).
MSA solutions consist of a “space diagram,” displaying the profiles (i.e., the reports), and a series of variable diagrams, called “item diagrams”. The configuration of the reports is the same in both types of displays. However, in an item diagram, a report is labelled in terms of its membership in one of the categories of the variable.
In general, the underlying hypothesis in an MSA is that the space can be partitioned by means of facets (Borg, 1992). More concrete, in this study, the space should be dividable

into sections by means of the elements within the facet for accuracy. The aim of applying MSA in this study is to show how well each variable can distinguish qualitatively good evaluation reports from qualitatively bad evaluation reports. In the item diagrams, there should be a clear distinction (indicated by a straight line) between reports that score high on this variable and reports that score low on this variable.

For the MSA, the *HOMALS* module of *SPSS 11* is chosen as a statistical package. Similarly to the MDS, an MSA gives information about the quality of the solution. In a HOMALS output for MSA, there are two important measures of quality: discrimination measures and eigen values. A discrimination measure in an MSA describes the ability of a variable to discriminate between the reports on one dimension: The higher the discrimination measure, the better the ability of this variable to discriminate between the reports in this dimension (see for example, Bühl & Zöfel, 2002). The eigen value is a measure of explained information whose calculation is based on averaging the discrimination measures for the respective dimension (Van de Geer, 1993).

In MSA, there is no *a priori* requirement as to the distribution characteristics of the variables nor their interrelationships. This method is also flexible in so far as it can deal with categorical as well as with ordinal data. For the present study, the data set with the ordinal data is chosen, because the analysis of these original, non-simplified data seems to be especially promising with regard to a detailed investigation.

There are several possible ways to conduct a concrete calculation. In the present study, the procedure should be systematic and replicable, so that another analyst using the same data set would arrive at the same set of variables. The main aim is to get the essential pattern from the most discriminating variables. The concrete steps undertaken to reach this aim will be described in relation to the realisation of the study.

2.2 Realisation of the study

2.2.1 Reports

The sample consists of 25 evaluation reports. These reports assessed different crime problems (Table 1), used different approaches (Table 2), came from different countries (Table 3) and were implemented in different locations (Table 4).

Table 1

Problem assessed

problem	Number
Drugs	3
Aggressiveness/violence	8
Vandalism/burglary	2
Prejudice/discrimination	5
Fear of crime	2
Treatment of victims	1
Crime in general	4

Table 2

Approach

approach	Number
Offender	13
Opportunity	1
Victim	3
Offender/opportunity	1
Offender/victim	3
Victim/opportunity	2
Offender/victim/opportunity	2

Table 3

Country of evaluation

Country	Number
Australia	1
Canada	1
Germany	11
Great Britain	2
Netherlands	1
Norway	2
USA	7

Table 4

Location of program

Location	Number
School	10
University	2
Neighbourhood	2
Public transport	2
Family/home	3
Clinic	3
Aid organisation	1
Police	1
Several locations	1

The majority of these reports is dated between 1990 and now. Additionally, some older reports were included that stand out either because they describe very typical crime prevention measures, or because they are very innovative.
A detailed list of the reports and their features can be found in Appendix D.

2.2.2 *Investigations*

In order to investigate the first hypothesis, the facet-theoretic model for evaluation was developed exactly as planned, and the variables of interest were explicated in a coding frame and a coding sheet. Using this coding sheet, the 25 evaluation reports were coded in a random order to avoid systematic influences of the order of coding. All reports were coded by the same person (the author). The reports were coded in a period of two weeks. It was made sure that no more than 6 reports were coded per day. On the basis of the results of coding, the analyses for testing the suitability of the evaluation criteria and the structure of the model were conducted as described in the Design.

The investigation of the second hypothesis centred on achieving the goal for the MSA formulated in the Design: The calculations should reveal the essential pattern from the most discriminating variables in a systematic and replicable way.
Before running the MSA, a discrepancy between the settings of HOMALS and the values in the data set had to be solved: For HOMALS, zero is a missing value, whereas in the data set zero is a possible value for each variable. Therefore, the data were linearly transformed. If there existed values of zero for a certain variable, a value of one was added to the original values, so that the original values were changed from 0 upwards to 1, from 1 upwards to 2 etc.
The MSA itself was realised in four steps: In the first step, a two-dimensional MSA was calculated with all variables. This calculation resulted in a table with discrimination measures on both dimensions for each variable. Only the most discriminating variables should be found. Therefore, variables with discrimination measures lower than 0.3 on both the first and the second dimension were excluded in the second step. In the third step, another MSA was calculated with the remaining well-discriminating variables. In the final step, it should be checked that no discriminating variables were excluded. For this reason, the other variables were included stepwise, in the order of their discrimination measures, starting with the highest value. Whenever a new variable was included, an item diagram for this variable was created. The diagram was partitioned into regions according to the scores on the variable, so that the reports in one region should have the same score on the respective variable. The

partitioning was indicated by a straight line. If the inclusion of a variable resulted in an item diagram for this variable with 3 or less wrong classifications, the variable was kept. Else, it was dropped and the proceeding continued with the next variable. This proceeding resulted in a final set including only the most discriminating variables.

The following example should clarify the concrete procedure of partitioning: As described above, some variables have more than one indicator. Accordingly, there are more than two possible values (see p.20).
The proceeding aimed at receiving the best possible solution, i.e., the solution with the least amount of errors. For this reason, the values were put together in two groups, so that the amount of errors was minimised. Say, for example, a variable had two indicators. In this case, there were three possible values: zero, which means that the expectations formulated in the indicators were not met; one, which means that the expectations for one indicator were met; and two, which means that the expectations for both indicators were met. In order to partition the item diagram for this variable into two regions, there were basically two possible ways. The researcher could either put together the values zero and one in one region and the value two in the other region. Or, the researcher could present the value zero in one region and the values one and two in the other region. The decision was taken on the basis of the amount of errors in the two possible solutions.

3. Results

3.1 A set of hierarchically structured mapping sentences

The mapping sentence on the most global level is an adaptation of the mapping sentence of Bilsky and Cairns (in press) to the situation of the evaluation of a report. The model in this study does not describe the evaluation of an intervention, but the evaluation of a report as the unit of analysis. Consequently, the intervention facet (A) was replaced by a facet for the report. As a report is the final product of an evaluation process, the meaning of the facet (B) for the timeframe of the evaluation changes and becomes a facet for the timeframe of the report.

		A:	*Report*		*B:*	*Status*
The evaluator (x) evaluates		(a1	project intervention)	,that is	(b1	published),
		(a2	program intervention)		(b2	unpublished)
	C:	*Standard*				
	(c1	utility)				
with respect to	(c2	feasibility)				
	(c3	propriety)				
	(c4	accuracy)				
	$R_{Evaluation}$					
→	(well done)	in terms of the standard (C4).				
	(not well done)					

Following the 11 accuracy standards in the Standards by the *Joint Committee*, the accuracy element was specified as follows:

		A:	*Report*		*B:*	*Status*
The evaluator (x) evaluates		(a1	project intervention)	,that is	(b1	published)
		(a2	program intervention)		(b2	unpublished)
	C4:	*Accuracy*				
	(c4.1	program documentation)				
	(c4.2	context analysis)				
	(c4.3	described purposes and procedures)				
	(c4.4	defensible information sources)				
with respect to	(c4.5	valid information)				
	(c4.6	reliable information)				
	(c4.7	systematic information)				
	(c4.8	analysis of quantitative information)				
	(c4.9	analysis of qualitative information)				
	(c4.10	justified conclusions)				
	(c4.11	impartial reporting)				
	$R_{Evaluation}$					
→	(well done)	in terms of the standard (C4).				
	(not well done)					

Note. The numbers in the facet C4 (Accuracy) indicate the placement of the elements in the *Standards* by the *Joint Committee*. Example: The notation “c4.5” shows that “valid information” is the fifth standard within the fourth section (Accuracy).

Based on key terms in the overview texts in *The Program Evaluation Standards* and complemented by Farrington (2003) the 11 standards were further elaborated. As explained in the Method section (p. 17), the fifth standard (valid information) was split into 3 elements (construct validity, internal validity and statistical conclusion validity) by referring to Farrington (2003).

		A:	*Report*		*B:*	*Status*
The evaluator (x) evaluates		(a1	project intervention)	,that is	(b1	published)
		(a2	program intervention)		(b2	unpublished)
	C4:	*Accuracy*				
	(c4.1.1	description of the unique features of the program)				
	(c4.1.2	description of the component parts)				
	(c4.1.3	description of the implementation of the program)				
	(c4.1.4	association of the components of the program with its effect)				
	(c4.2.1	geographic location of the program)				
	(c4.2.2	its timing)				
	(c4.2.3	the political and social climate surrounding it)				
	(c4.2.4	the staff)				
	(c4.2.5	pertinent economic conditions)				
	(c4.3.1	the evaluation purposes)				
	(c4.3.2	description of the procedures)				
	(c4.4.1	description of the sources of information)				
	(c4.4.2	variety of sources)				
	(c4.4.3	description of the sample)				
	(c4.4.4	dealing with missing data)				
with respect to	(c4.5.1	construct validity)				
	(c4.5.2	internal validity)				
	(c4.5.3	statistical conclusion validity)				
	(c4.6.1	assessment of the reliability of the instruments)				
	(c4.7.1	assuring that all information is as free from error as is possible and kept secure)				
	(c4.8.1	use of initial exploratory analyses)				
	(c4.8.2	use of more sophisticated and complex analyses)				
	(c4.8.3	visual displays)				
	(c4.9.1	set of categories)				
	(c4.9.2	test of categories for validity and reliability)				
	(c4.9.3	meaningfulness of conclusions and recommendations)				
	(c4.10.1	adequate interpretation of statistics)				
	(c4.10.2	relation of conclusions to statistical results)				
	(c4.10.3	possible alternative explanations for results)				
	(c4.11.1	neutral and objective style of reporting at any stage of the report)				

→ $R_{Evaluation}$ (well done / not well done) in terms of the standard (C4).

Note. Again, the numbers in the facet C4 (Accuracy) indicate the assignment of the elements to standards within the work of the *Joint Committee*. Example: The notation “c4.9.1” shows that “set of categories” is the first specification of the ninth standard within the fourth section (Accuracy).

3.2 Data collection – the coding frame and its application

For the purposes of the coding frame, the 30 elements within the accuracy facet in the most specified mapping sentence were numbered from v1 to v30. They were measured by the indicators that are presented in Table 5.

As explained in the Method, the variables are measured as the sum of the respective indicators. Consequently, the value assigned to a certain report corresponds to the number of indicators which the expectations are met for. The variable v16, for example, has four indicators, which means that values between 0 and 4 can be assigned to a certain report for this variable.

Table 5

Variables and indicators

	Label	Indicators
v1	Description of the unique features of the program	• Mention of basic assumptions/theoretical background • Mention of aims of the program
v2	Description of the component parts	• Description of the content of the program (e.g., sessions, examples)
v3	Description of the implementation of the program	• Comparison between design and implementation
v4	Association of the components of the program with its effect	• Reference to intended or real effect of the program
v5	Location of the program	• Mention of geographic location • Description of setting
v6	Its timing	• Mention of date of implementation • Description of time schedule
v7	The political and social climate surrounding it	• Reference to political supports or critics • Basic facts about the problem
v8	The staff	• Mention of staff and statement on qualification and motivation of the staff
v9	Pertinent economic conditions	• Reference to time and/or money
v10	Evaluation purposes	• Mention of objectives or hypotheses • Mention of intended use of results
v11	Description of the procedures	• Description of the process of data collection • Description of data analysis
v12	Description of the sources of information	• Description of the instruments or way of measurement • Description of the quality of the sources
v13	Variety of sources	• At least one variable was assessed by different sources to allow for data loss

v14	Description of the sample	• Description of the formal procedure of drawing the sample • Description of the sample/the unit of analysis • Information about changes in the sample
v15	Dealing with missing data	• Information about dealing with missing data
v16	Construct validity	• Adequacy of the operational definition • Validity of the instrument • Multiple sources of information • Assessment of unintended effects
v17	Internal validity	• Experimental manipulation • Control group • Random assignment • At least two times of measurement (Pre-Post) • Blind participants and blind observers • Assessment of other possible influences or mediators
v18	Statistical conclusion validity	• The data do not violate the underlying assumptions of the statistical test(s) • Calculation of effect sizes or equivalents
v19	Assessment of the reliability of the instruments	• Assessment of the reliability of the instrument measuring the central construct
v20	Assuring that all information is as free from error as is possible and kept secure	• Mention of attempts to find and correct errors
v21	Use of initial exploratory (descriptive) analyses	• Use of analyses that assess the nature or the quality of the data
v22	Use of more sophisticated and complex analyses	• Mention and description of the sophisticated procedure • Giving appropriate statistics
v23	Visual displays	• The report contains at least one visual display
v24	Set of categories/approach towards structuring the qualitative data	• Existence of a set of categories to structure qualitative information
v25	Assessment of the quality of categories/structure	• Assessment of the quality of the categories
v26	Meaningfulness of conclusions and recommendations	• Reference to the qualitative information in the conclusions or recommendations
v27	Adequate interpretation of statistics	• The values of the important statistics have been interpreted
v28	Relation of conclusions to statistical results	• Relation of the main conclusions to the statistical results
v29	Possible alternative explanation for results	• Mention of possible alternative explanations for the results
v30	Neutral and objective style of reporting at any stage of the report	• Neutral language, evaluative words only in the introduction, the interpretation, or the discussion

For detailed information on the coding frame, see Appendix B.

In the course of the coding process, it was found out that one of the indicators was not applicable to the evaluation reports, namely the first indicator of variable 18: “The data do not violate the underlying assumptions of the statistical test(s).” The author found that the appropriate information was not included in the present reports.

The results of coding and the results of the process of dichotomising are presented in Appendix C.

3.3 Structure of the model

3.3.1 Suitability of the evaluation criteria for further analyses – Results of descriptive analyses

Frequency data of the ordinal and dichotomised variables are presented in Table 6 (a) and Table (b).
As can be seen in Table 6 (b), 10 variables were excluded as the distributions of frequencies were too skew and, hence, did not meet the 20%/80% criterion. The coding of these variables resulted either in less than 5 cases in the 0 category (almost all reports were qualitatively good in terms of this variable) or in the 1 category, respectively (hardly any report was qualitatively good in terms of this variable).
These 10 variables were excluded from all further analyses, therefore.

Table 6 (a)

Frequencies of reports per coding category – ordinal variables

v	Frequencies					
	0	1	2	3	4	5
v1	0	6	19	-	-	-
v2	3	22	-	-	-	-
v3	8	17	-	-	-	-
v4	5	20	-	-	-	-
v5	1	12	12	-	-	-
v6	2	13	10	-	-	-
v7	5	7	13	-	-	-
v8	23	2	-	-	-	-
v9	14	11	-	-	-	-
v10	2	17	6	-	-	-
v11	0	3	22	-	-	-
v12	4	14	7	-	-	-
v13	23	2	-	-	-	-
v14	3	4	10	8	-	-
v15	24	1	-	-	-	-
v16	2	8	12	2	1	-
v17	2	4	3	5	8	3
v18	0	23	2	-	-	-
v19	20	5	-	-	-	-
v20	13	12	-	-	-	-
v21	1	24	-	-	-	-
v22	8	6	11	-	-	-
v23	3	22	-	-	-	-
v24	16	9	-	-	-	-
v25	23	2	-	-	-	-
v26	12	13	-	-	-	-
v27	2	23	-	-	-	-
v28	6	19	-	-	-	-
v29	19	6	-	-	-	-
v30	7	18	-	-	-	-

Table 6 (b)

Frequencies of reports per coding category – dichotomous variables

v	Frequencies	
	0	1
v1	6	19
v2[a]	3	22
v3	8	17
v4	5	20
v5	13	12
v6	15	10
v7	12	13
v8[a]	23	2
v9	14	11
v10	19	6
v11[a]	3	22
v12	18	7
v13[a]	23	2
v14	17	8
v15[a]	24	1
v16	10	15
v17	14	11
v18[a]	23	2
v19	20	5
v20	13	12
v21[a]	1	24
v22	14	11
v23[a]	3	22
v24	16	9
v25[a]	23	2
v26	12	13
v27[a]	2	23
v28	6	19
v29	19	6
v30	7	18

[a] These variables did not meet the variation criterion (80/20) and were excluded, therefore.

3.3.2 Structure of the model – Results of Multidimensional Scaling (MDS)

An MDS was calculated with both the ordinal and the dichotomous variables. The outputs (Figures 6 & 7) show that some variables in the middle are placed closer together, which means that their intercorrelations are higher (Guttman, 1968).

Figure 6: MDS configuration of the dichotomous variables

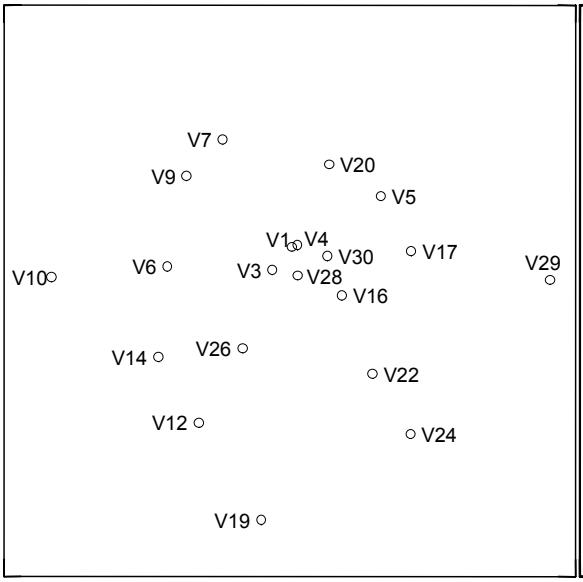
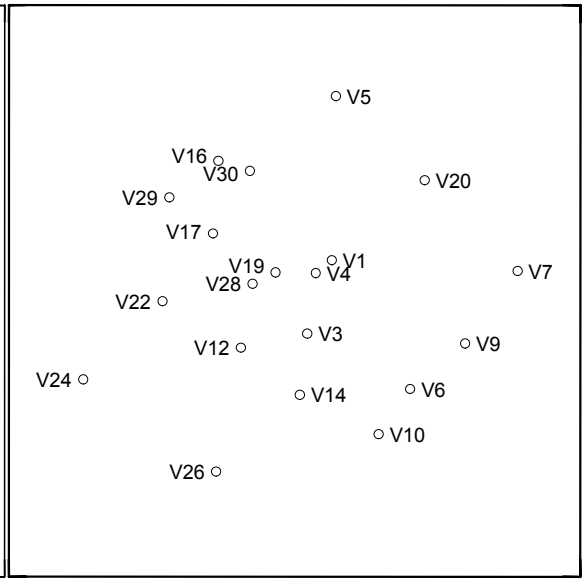


Figure 7: MDS configuration of the ordinal variables



As expected, the variables are arranged in a circular order, and the space can be partitioned into wedge-shaped regions (Figure 8). The program (v1) is placed in the centre (visualised by a dot). It is surrounded by variables that describe different aspects of the evaluation process.

To visualise the structure of the variables, the graph can be partitioned into regions, so that each region contains only variables with a common feature (see for example: Borg & Staufenbiel, 1993). In this way, the space can be partitioned into 8 regions within which the variables form a homogeneous set (Figure 8). Each region deals with one qualitatively different aspect, varying from scientific criteria (validity, objectivity, reliability) and different types of analyses (quantitative, qualitative) to the implementation, the context and aspects of the proceeding (avoiding errors and assessing the need for the program).

Figure 8 : Configuration of regions in the MDS output for variables that are measured on an ordinal scale – a polar structure

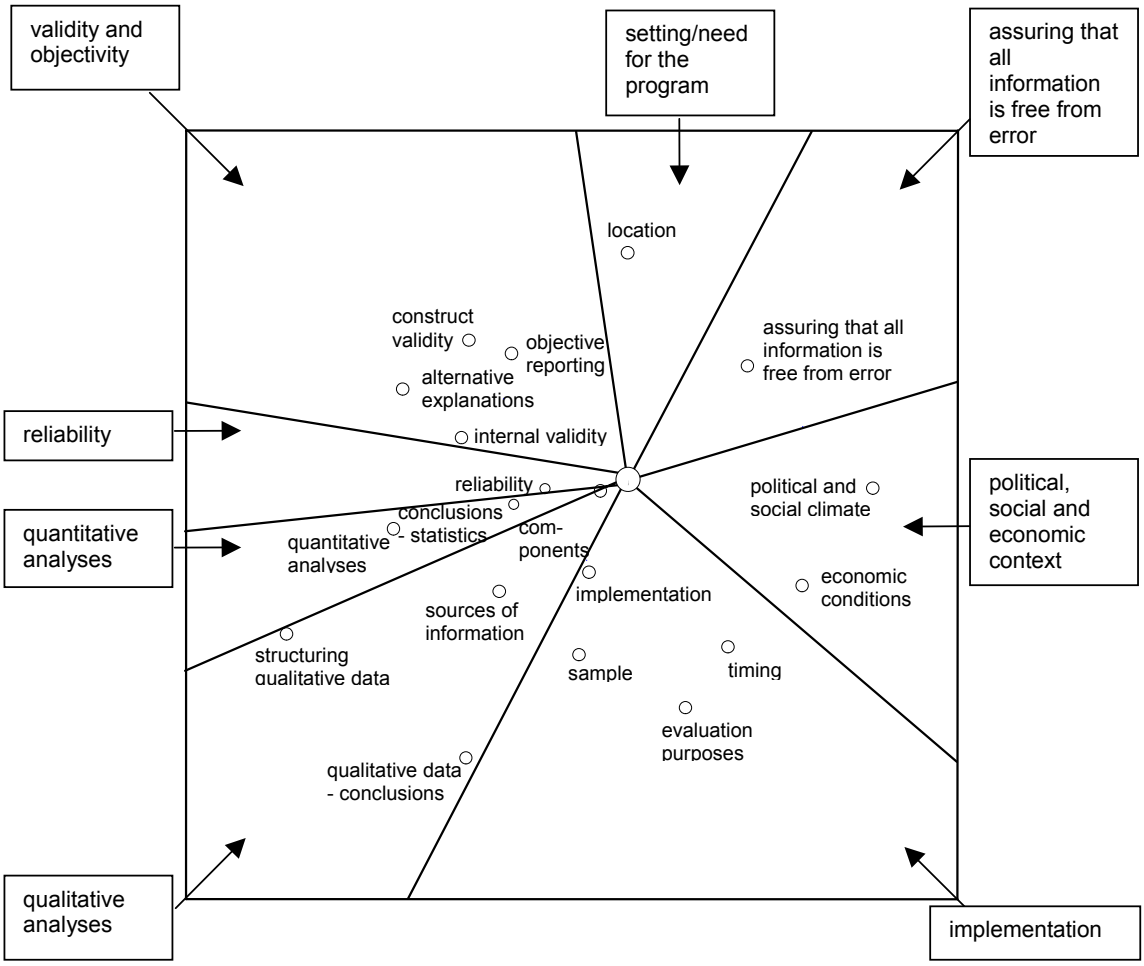
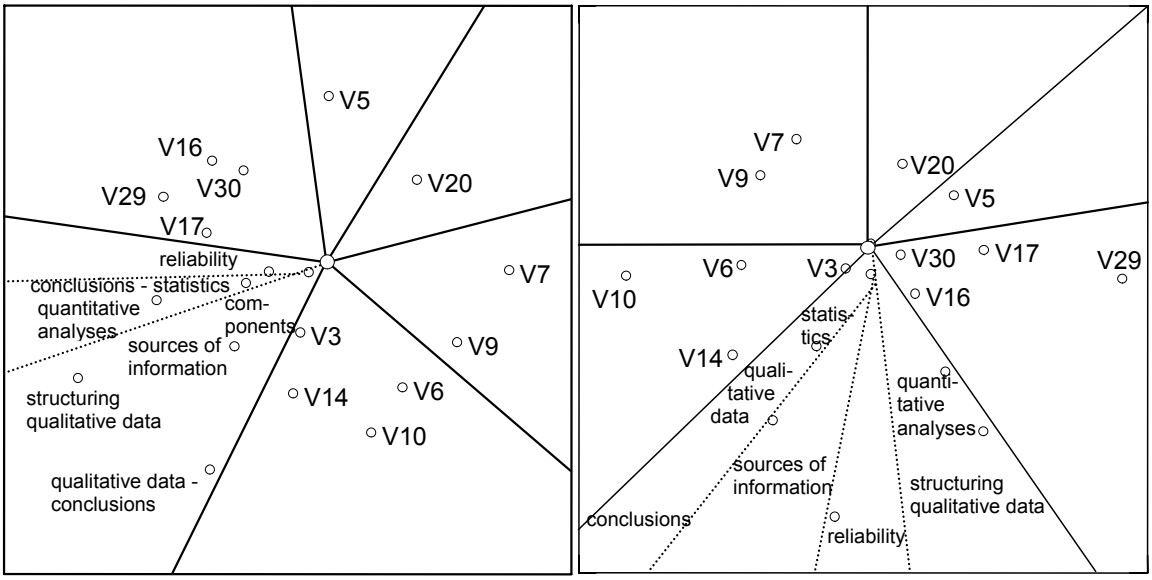


Figure 9 : Configuration of regions in the MDS output – a comparison between ordinal and dichotomous data

(a) Ordinally scaled data

(b) Dichotomous data



Note. The program is placed in the centre of the graph and is symbolized by a dot (◯). In figure 9 (b), the dot representing the program hides v4 (components), which is positioned at almost the same place.

The general structure was confirmed in the MDS of the dichotomous data (Figure 9). The results differ in the structure with respect to those variables that deal with analyses, namely:

- V12: Description of the sources of information (sources of information)
- V22: Use of more sophisticated and complex analyses (quantitative analyses)
- V24: Set of categories/approach towards structuring the qualitative data (structuring qualitative data)
- V26: Meaningfulness of conclusions and recommendations (qualitative data-conclusions).

Additionally, V19 (Assessment of the reliability of the instrument, reliability) is displaced.

For the ordinally scaled data, the MDS presented these four variables in two regions, one for quantitative and one for qualitative analyses (Figure 9 (a)). In contrast, the MDS for the dichotomous data revealed 3 regions (Figure 9 (b)): The first partition contains variables for structuring data (quantitative analyses, structuring data). The second region contains the variable “sources of information”, and the third region contains variables that relate results of analyses to conclusions (statistics -, qualitative analyses – conclusions).

The coefficient of alienation is K=0.27 for the solution for the ordinal variables and K=0.23 for the solution for the dichotomous variables. That means, the alienation coefficients in this study are higher than Louis Guttman’s recommendation. It should be taken into account, however, that 20 variables are displayed in an only two-dimensional space (see Method, p.22).

3.4 Usefulness of the model in assessing the quality of evaluation reports

The calculation of the MSA resulted in a table with discrimination measures on both dimensions for each variable. As can be seen in Table 7, the discrimination measures are generally higher on the first dimension. Eleven variables were included in the initial set for calculation, because their discrimination measures were greater than 0.3 (Table 7). Then, two of the initially dropped variables were re-included (for an explanation of the proceeding, see Method section).

Table 7

Discrimination measures

Name of variable		Dimension	
		1	2
V1[b]	Description of the unique features of the program	.533	.099
V3[b]	Description of the implementation of the program	.514	.049
V4[b]	Association of the components of the program with its effect	.366	.037
V5	Location of the program	.147	.061
V6	Its timing	.201	.070
V7[b]	The political and social climate surrounding it	.041	.418
V9	Pertinent economic conditions	.071	.262
V10	The evaluation purposes	.175	.260
V12[b]	Description of the sources of information	.400	.356
V14[b]	Description of the sample	.383	.335
V16[b]	Construct validity	.353	.454
V17[b]	Internal validity	.702	.141
V19	Assessment of reliability of the instrument	.240	.003
V20	Assuring that all information is as free from error as is possible and kept secure	.051	.155
V22[b]	Use of more sophisticated and complex analyses	.661	.314
V24	Set of categories/approach towards structuring the qualitative data	.015	.185
V26	Meaningfulness of conclusions and recommendations	.111	.131
V28[b]	Relation of conclusions to statistical results	.690	.009
V29	Possible alternative explanations for results	.120	.023
V30[b]	Neutral and objective style of reporting at any stage of the report	.358	.030

[b] Variables with discrimination measures greater than .3 were included in the first set.

The final solution has an eigen value (Bühl & Zöfel, 2002) of .42 for the first dimension and .22 for the second dimension .

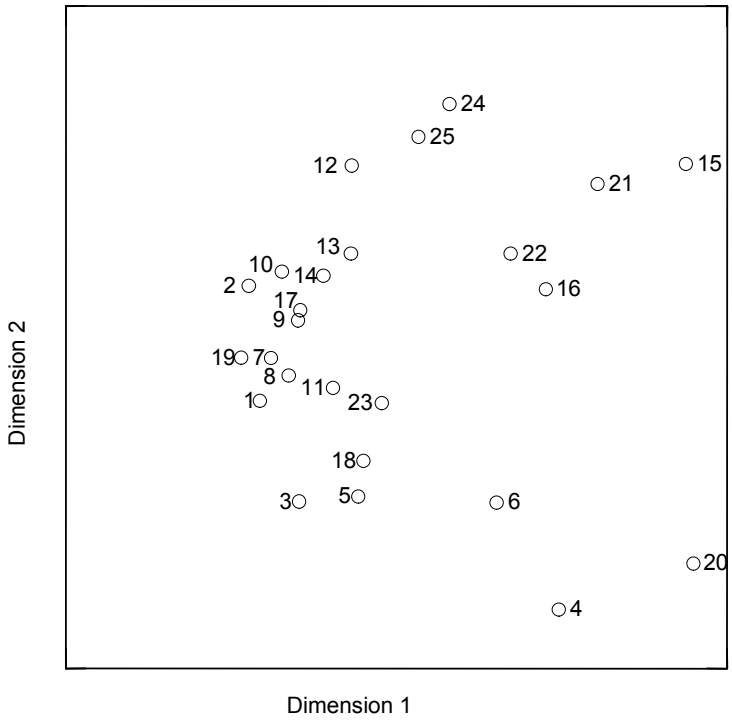
It shows that following 13 variables could distinguish between the reports:

- V1 Description of the unique features of the program
- V3 Description of the implementation of the program
- V4 Association of the components of the program with its effect
- V7 The political and social climate surrounding it
- V9 Pertinent economic conditions
- V12 Description of the sources of information
- V14 Description of the sample
- V16 Construct validity

- V17 Internal validity
- V19 Assessment of the reliability of the instruments
- V22 Use of more sophisticated and complex analyses
- V28 Relation of conclusions to statistical results
- V30 Neutral and objective style of reporting at any stage of the report.

Based on these 13 variables, a two-dimensional space diagram of the evaluation reports was produced (Figure 10). A large cluster of reports can be found in the left part of the diagram.

Figure 10: Space diagram of evaluation reports



Note. Numbers correspond to the numbering of the reports in the Appendix.

Figure 11 presents the second graphical output of an MSA: the item diagrams. A straight line in each diagram separates one category (or a group of categories) from the other. As explained in the Method (p.29), this separation represents the solution that minimises the amount of wrong classifications.

The orientation of the lines is vertical or diagonal in most cases. Only the lines for v7 and v9 are completely horizontal. This orientation, together with the positioning of these lines leads to the effect that v7 and v9 are those variables with the best discrimination within the cluster of reports mentioned above. These variables seem to play a special role for the assessment of the quality of the 25 reports.

Figure 11: Item diagrams

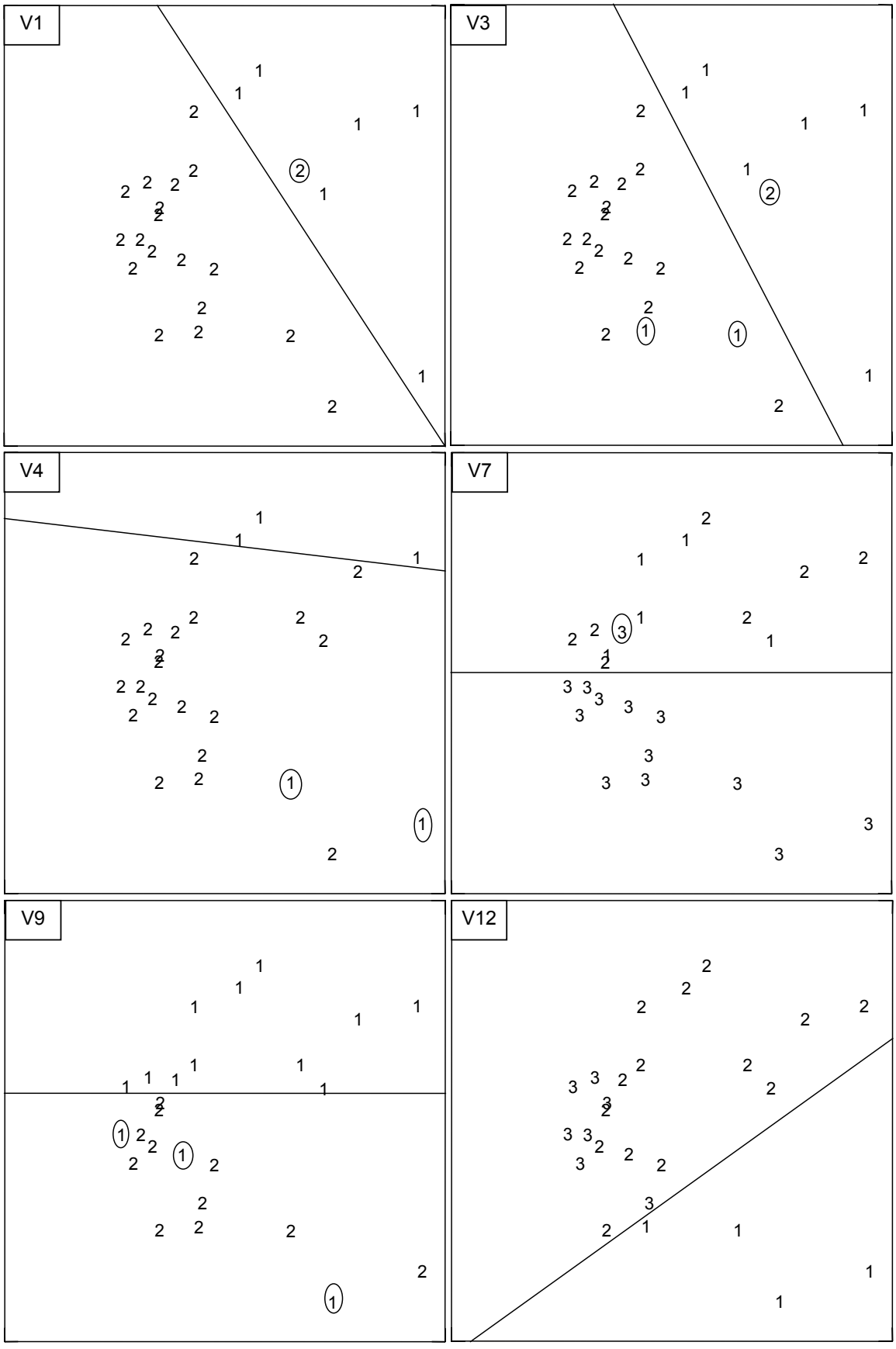


Figure 11 – continued

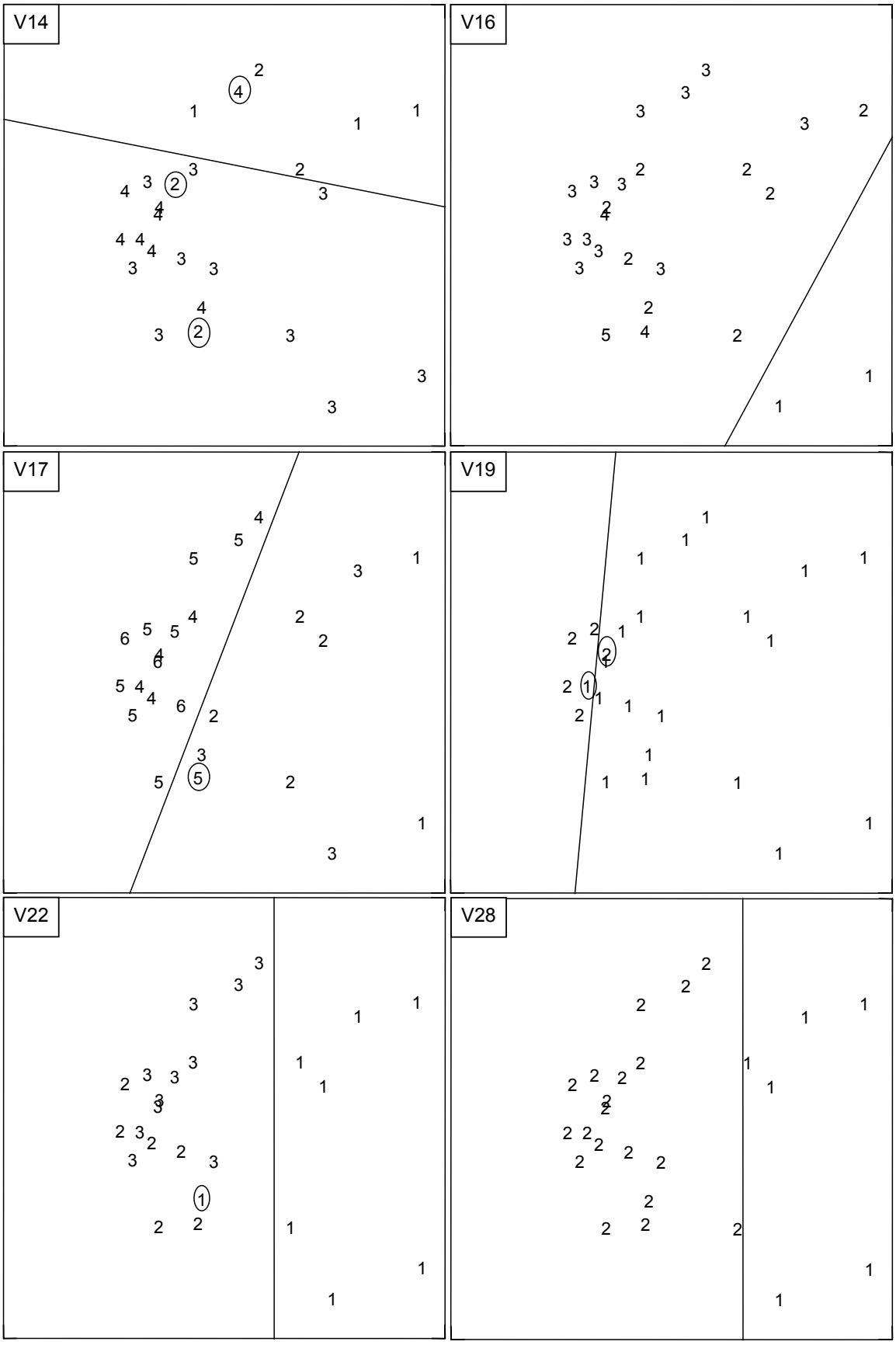
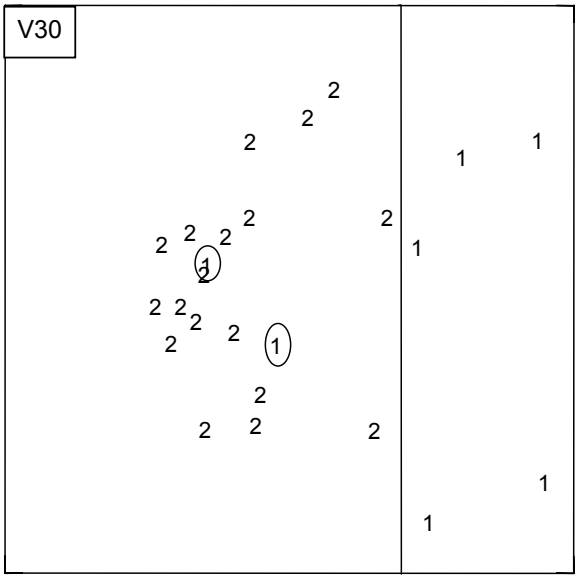


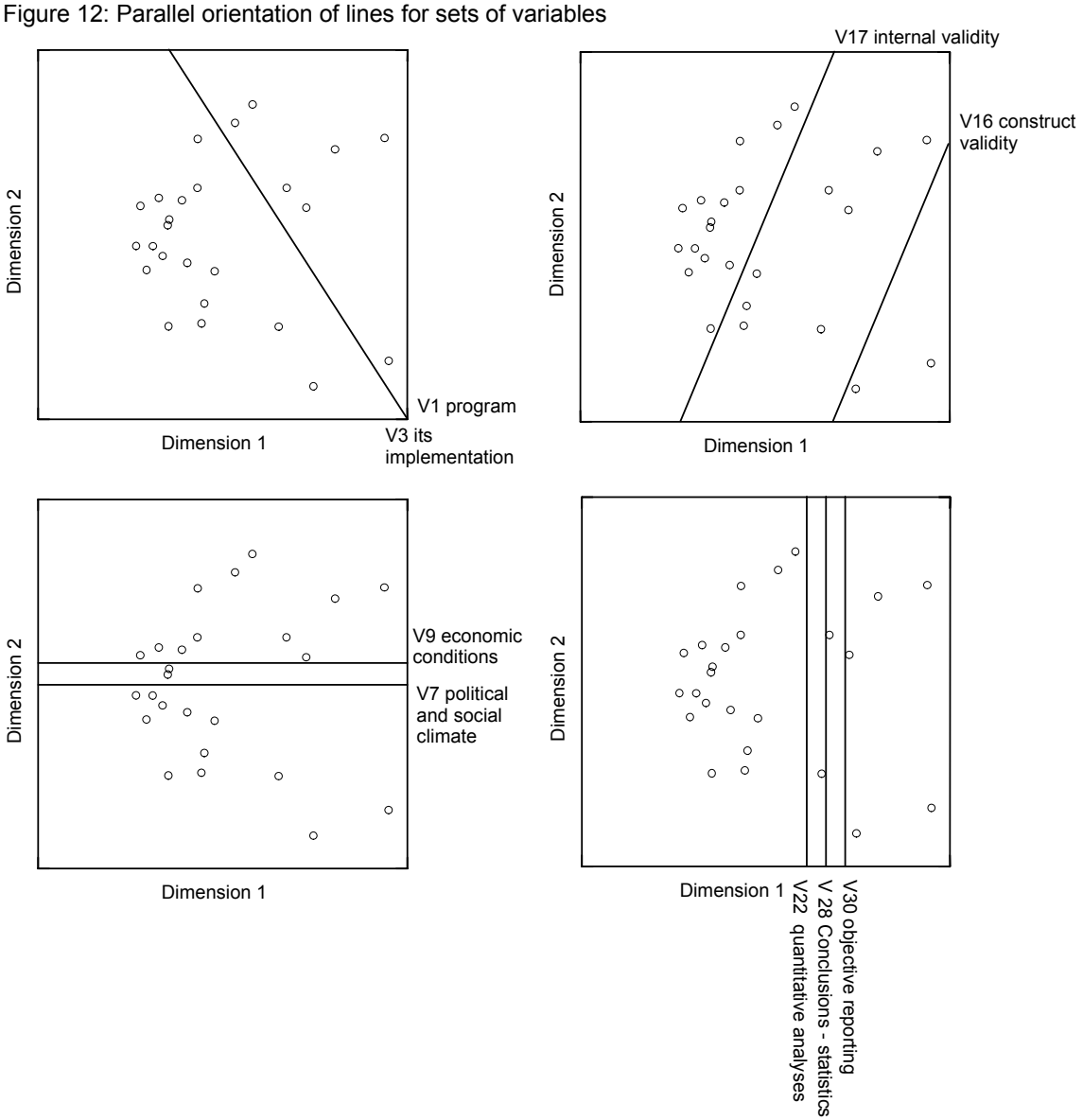
Figure 11 – continued



Note. Numbers represent the membership of each report in the categories the variable. Errors are marked by circles. For data-analytic reasons, the data were linearly transformed before running the MSA (see Method, p.27).

Some variables distinguish in a similar way (Figure 12), indicated by a parallel orientation of the accordant lines. This means that there is the same underlying proceeding in the assessment of the quality of the reports for the program and its components (v1 and v3), for different aspects of validity (v16 and v17), for aspects of the practical context (v7 and v9) and for quantitative analyses and objectivity (v22, v28, and v30) respectively.

Figure 12: Parallel orientation of lines for sets of variables



4. Discussion

In the present study, it was investigated whether facet theory can make a contribution to a new area of application, which is in the focus of current research and political discussions: crime prevention. Facet theory was used to systematise evaluation research and to integrate concrete evaluation studies into a common facet theoretic framework. The implications of the results of this research will be evaluated and interpreted with respect to the original hypotheses. In addition, the discussion offers a synopsis of the results, in order to facilitate the development of new ideas and to highlight possible directions for future research.

4.1 Confirmation of hypotheses

Both hypotheses formulated in the Introduction were confirmed, as will be clarified in the following sections.

4.1.1 Confirmation of hypothesis 1

The first hypothesis was: It is possible to state evaluation research actions in terms of facet theory to reach conceptual clarity. This hypothesis was investigated by developing and testing a facet-theoretic model for evaluation.

Development of a set of hierarchically structured mapping sentences

Program evaluation standards had not been thoroughly studied within a facet framework so far. For the present study, the only existing facet-theoretic model for evaluation standards (Bilsky and Cairns, in press) was specified step by step. The facet that covers evaluation standards was in the centre of the present research. The concretisation of this facet was based on what was thought to be the most important international literature on standards for evaluation. Substance was given to this first model by using the Standards by the Joint Committee (1994), Rossi et al. (2002) and the standards for evaluating crime prevention measures by the Campbell Collaboration (Farrington, 2003). In this way, important standards from different sources could be integrated in a single model. Information from different sources was accumulated, and the existing literature was summarised in 30 variables. Because of the significance of the chosen literature, the model itself should contain the central evaluation standards for programs today.

Empirical test of the model

Applicability of the coding frame

The model was applied to a heterogeneous, however, non-representative sample of evaluation reports of crime prevention. In this process, the coding frame as the instrument for data collection served as link between the abstract model and concrete evaluation reports. In spite of the heterogeneity in the sample, the coding frame was applicable to all reports. So, it seems to be general, i.e. it goes far beyond specific programs, problems, locations, and national specifics.

Suitability of the evaluation criteria for further analyses

The suitability of the evaluation criteria, i.e. their usefulness for giving information on differences between reports, was formulated as a prerequisite for the inclusion of these criteria in further analyses (see Method). Descriptive analyses gave information on the suitability of the chosen evaluation criteria. The results showed that the majority of the criteria could distinguish between the reports. However, the distributions for 10 of the criteria were too skewed according to the criterion formulated in the Method section. Generally spoken, there were two cases:

1. For some variables, almost all reports were thought to be “well done” in terms of this standard. These variables exclusively cover descriptive variables: content of the program, description of the procedures, use of initial exploratory (descriptive) analyses, visual displays.
2. For other variables, almost no report was thought to be “well done” in terms of this standard. All of these variables deal with some kind of methodological procedure: assessment of the qualification and motivation of the staff, dealing with missing data, variety of sources to allow for data loss, statistical conclusion validity, assessment of quality of categories/structure.

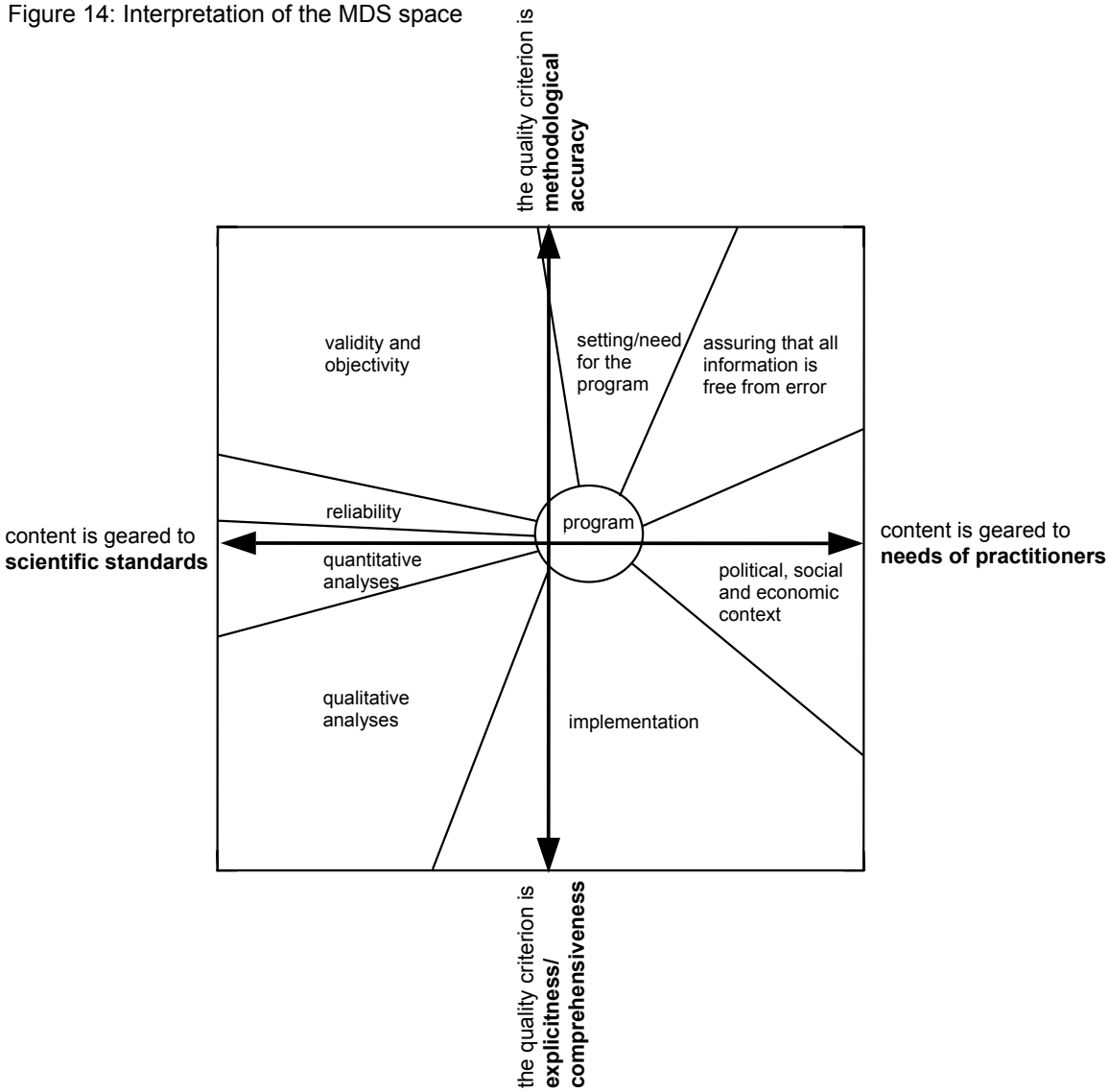
This is to say that the application of some descriptive variables to the sample under study did not make sense, because the level of expectation that is expressed in these variables is very low. In contrast, the application of some variables dealing with methodological procedures did not make sense, because the level of expectation that is expressed in these variables is very high. The finding that certain types of evaluation criteria make high demands on reports can be regarded as a first hint. However, the reader should keep in mind that this is only valid for the reports under investigation in the present research.

Structure: Revealing the relationships between evaluation criteria by Multidimensional Scaling (MDS)

The analysis of the structure of the variables concentrated on revealing similarities and dissimilarities between them. As expected, there is a circular order among variables, and the MDS space can be partitioned into wedge shaped regions. The program is placed in the centre and is surrounded by variables that deal with the evaluation process. By partitioning the surrounding into homogeneous regions, the results give further insight into the exact structure. The grouping of variables in homogeneous regions resulted in a meaningful structure. In this way, the application of facet-theory led to conceptual clarity, which means that the first hypothesis was confirmed.

Figure 14 goes beyond the partitioning in the Results and offers an interesting interpretation of the space:
Moving from the very left part of the diagram to the very right part, the content that is assessed by the variables changes. While the content of the variables on the left is geared to scientific standards (e.g., validity), the content of the variables on the right is geared to needs of practitioners. Practitioners want to know, for example, about the costs and the financial support (economic context). This distinction can be traced back to the historical dispute between researchers who emphasise scientific methods (e.g., Campbell and Scriven) on the one hand and researchers who emphasise the importance of the social, political and economic context (e.g., Weiss and Wholey) (see Introduction, p.2).
Moving from the top of the diagram to the bottom, the quality criterion changes from methodological accuracy (e.g. avoiding errors) to explicitness/comprehensiveness (e.g., description of the implementation).
Consequently, each variable can be described as a combination of content and criterion. This structure, which was revealed by the MDS, will be elaborated in a suggestion for future directions.

Figure 14: Interpretation of the MDS space



4.1.2 Confirmation of hypothesis 2

The second hypothesis was: Within a facet-theoretic framework, the quality of evaluation reports can be assessed, so that a distinction can be made between “well-done” and not “well-done” evaluations. The central aspect under study was the usefulness of the chosen evaluation criteria in assessing the quality of reports.

Assessing the quality of reports by Multiple Scalogram Analysis (MSA)

The results showed that the mapping sentence in this study is promising for assessing the quality of evaluation reports. Facet-theoretic analyses illustrated that the quality of the evaluation reports under study could be assessed by criteria formulated in the model. The

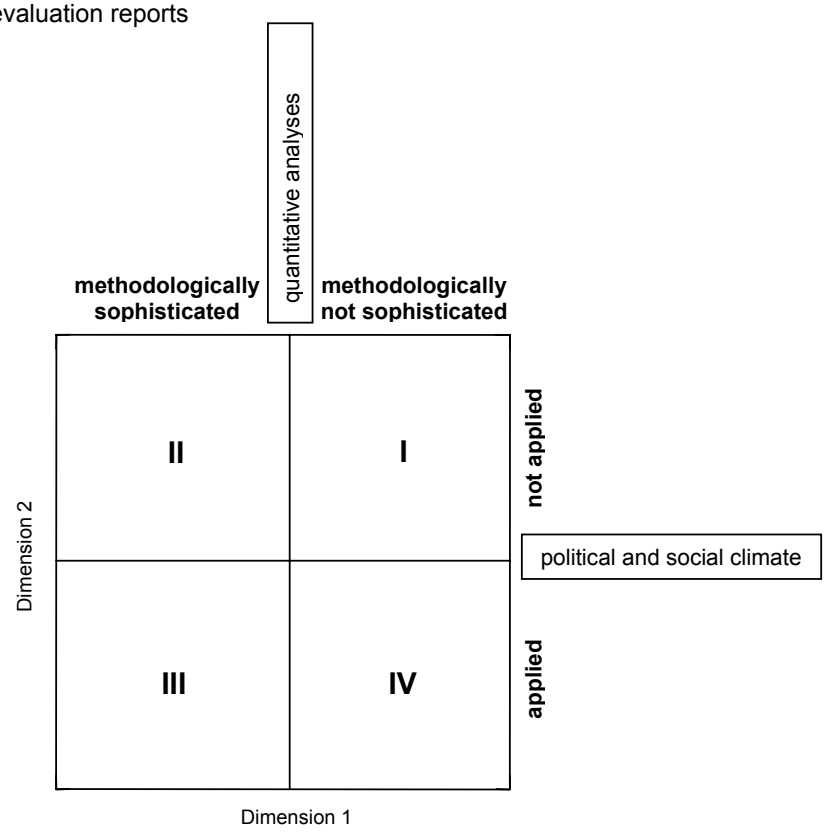
MSA created displays of the structure of the reports that visualised their quality. By referring to these displays, the analyses demonstrated the potential of using certain criteria for distinguishing between reports with respect to their quality.

Some of those variables that are useful for assessing the quality of reports, played a special role in the Results:
The Results shed light on the importance of the political, social context and economic context of an evaluation. The corresponding variables were the only variables that partitioned the space into regions by drawing a *horizontal* line (v7, v9; see Figure 11 & 12). Their importance was confirmed by the observation that these were the variables with the best potential for distinguishing within the big cluster of reports.
In spite of their importance for assessing the quality of reports, the results of the MSA showed that this type of variables was underrepresented in the total set. Only the variables for social, political and economic context have high discrimination measures exclusively on the second dimension. All other variables separate by a either vertical or diagonal line.
While the variables for political, social and economic context separate by a *horizontal* line, it was shown that three variables separate by a *vertical* line: use of more sophisticated and complex analyses (v22), relation of conclusions to statistical results (v28), and neutral, and objective style of reporting (v30) (see Figure 12). A combination of these horizontal and vertical lines can be used in order to describe the reports by a typology (Figure 15). For illustrative purposes, one variable was chosen from each set (namely v7 from the first set and v22 from the second set). In this typology, a report can either be methodologically sophisticated or not. At the same time, it can be either applied or not applied.

This synopsis of the MSA results provides ideas for the improvement of certain types of evaluation reports. Figure 16 shows which criteria are important within each quadrant. The underlying question is: How can a certain type of report be improved?
In every quadrant, at least one of those variables which cover a description of the program or the evaluation (v1, v3, v12, v14) is important. This means that every report in this study can benefit from an explicit and comprehensive description. Interestingly, for the least sophisticated reports in terms of the model (quadrant I), the results emphasise an additional aspect: Apart from the pure existence of a description, the way of describing can be improved by using an objective style of writing (v30).
It is not surprising that reports which are methodologically sophisticated but not applied (quadrant II) can be improved by a description of one aspect of the practical context, namely the economic context (v9). In contrast, reports that are applied but not methodologically sophisticated (quadrant IV) can benefit from the elaboration of methods (v28 relation of

conclusions to statistical results) and the application of methodological criteria (v16 construct validity).

Figure 15: A typology of evaluation reports



The best reports can be found in the very left segment of the third quadrant. A short description of the three “best” reports in terms of the model shows, that the facet-theoretic model and the corresponding coding were able to deal with a broad bandwidth of contents. Obviously, the application of the model leads to a balanced evaluation of different problems, approaches, and methods. It seems to be open towards different methods, it takes into account that there are different ways to conduct a good evaluation. So, what are especially good evaluations in terms of the model?

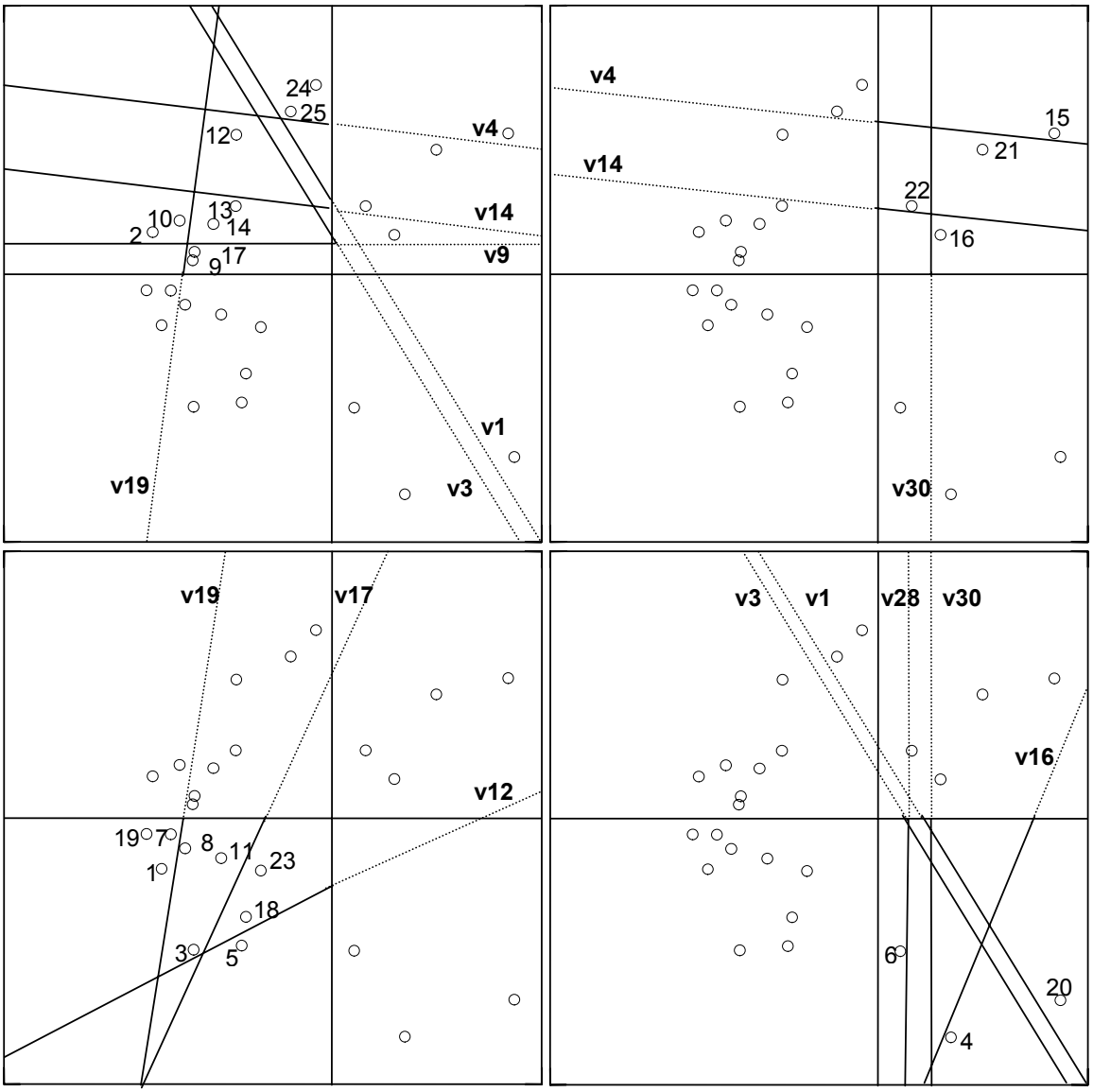
Report 1: Botvin, G. J. (1997). School-based drug abuse prevention with inner-city minority youth. *Journal of child and adolescent substance abuse*, 6(1), 5-20.

This evaluation of a drug prevention program was conducted in the United States of America. It assesses the value of a skills training for school children to resist social influences. In a quasi-experimental design, questionnaires and physiological measures were used as data sources. The results were analysed in a *General Linear Model (GLM),* allowing for the integration of different statistical methods.

Report 7: Farley, D., & Magill, J. (1988). An Evaluation of a group program for men who batter. In G. S. Getzel (Ed.), *Violence. Prevention and treatment in groups* (pp. 53-65). New York/London.
This Canadian report describes the evaluation of a violence prevention and treatment program for men who batter at home. In a group, men should learn alternative behaviours. Clinical file notes, a scale of social functioning and qualitative data (impressions and observations of the group leader) served as data sources in Pre- and Post-Test. The analyses integrated quantitative and qualitative data.
Report 19: Voß, H.-G. W. (n.d.). *Professionalität in der polizeilichen Arbeit mit Opfern und Zeugen*. Unpublished manuscript.
This German report evaluates a training for police officers that should lead to a professional way of dealing with victims and witnesses. For that purpose, the program was implemented in a quasi-experimental design, and its impact was tested by letting victims and witnesses fill in a questionnaire. The analyses are based on correlations and comparisons of means.
These three examples demonstrate that keeping to the standards does not automatically lead to a decrease in flexibility. It is encouraging, that some reports met almost all criteria of the model. Nevertheless, it has to be noted that no report was up to all demands formulated in the 30 variables.

In the present analysis, seven of the remaining 20 variables did not distinguish between the reports. These variables cover a broad range: specific features of the implementation, evaluation purposes, avoiding errors, qualitative data analyses, conclusions and interpretations. Because of their central role in the literature on program evaluation, these variables should not be excluded in future analyses. Instead, effort should be made to refine and clarify the structure among the variables. A refinement of the structure will be discussed in relation to an elaboration of the model.

Figure 16: Regions within quadrants



4.2 Limitations

It should also be noted, however, that this study has several limitations. Methodological limitations include small sample size and lack of controls in the coding process. All reports were rated by the researcher who had developed the coding frame. Because of limited resources, no independent rater coded the reports. Therefore, it was impossible to account for errors and biases. For future research, it would be desirable to recruit a second, independent rater and to calculate interrater-reliability as a measure of quality for the coding frame.

4.3 Practical relevance

In the last years, there has been a growing need for improving the evaluation of crime prevention measures. Various attempts have been made to improve the process of evaluation. The interests of practitioners always played an important role in this context. This issue was especially important with regard to a balance between the concerns of practitioners contrasted to those of scientists. Rossi (Chen & Rossi, 1980) took into account this diversity of interests in his “multi-goal” approach and in his textbook (Rossi et al., 2002) (see Introduction). In the same way, a model for evaluation should attempt to integrate different, sometimes conflicting points of view. That means, it should cover variables for methodological accuracy as well as variables that are very practical, such as costs, time and other resources .

In this context, guidelines for practitioners play a central role. They are intended to serve as link between scientific standards on the one hand and the concerns of practitioners on the other hand. They should make a contribution to the improvement of the quality of evaluation. Recently, Hupfeld (2004) has developed such guidelines for practitioners. They cover quality management in planning, implementing and evaluating crime prevention projects. The guidelines were elaborated in cooperation with practitioners as well as methodological experts. To find out crucial aspects of a good evaluation, practitioners and methodological experts were asked to judge the relevance of several steps in an evaluation. Table 8 shows those aspects that were rated as essential. The right column presents the corresponding aspects in the model which was developed in the present study.

Important questions are the following: Do these practical guidelines contain the same aspects as the model in the present study? Does the model account for interests of practitioners? Many aspects that Hupfeld mentions are part of the model. No variable that can be assigned to *Accuracy* is disregarded. Furthermore, the model in the present study accounts for many additional variables. Most of them deal with methods, analyses and scientific criteria of quality.
Nevertheless, the comparison reveals that Hupfeld gives top priority to a comprehensive explanation of the program and to aspects of the practical context. While in the present model the program (v1) and the political and social context (v7) are covered by only two variables, Hupfeld formulates several essential questions for each of these two aspects (Table 8). Obviously, the rationale of the program as well as practical aspects were regarded as essential by experts in the field of evaluation, and consequently, should play a more important role in a model for evaluation.

Table 8

Hupfeld	**Döring**
Are the main aims really related to crime prevention?	Precondition for inclusion of reports in my analysis
Are the main aims mentioned?	v1:Mention of aims of the program
Do the subordinate aims really serve the main aim?	v1: Mention of aims of the program
What measures should be conducted?	v1: Description of the unique features of the program v2: Description of the component parts
Is the submitted time span for the project long enough to achieve the aims?	v2: Description of the content of the program (e.g., sessions, examples) v6: time schedule
What is the expected impact?	v4: Association of the components of the program with its effect
What is the problem?	v7: Basic facts about the problem
Is there evidence for the problem?	v7: Basic facts about the problem
Who is already concerned with the problem?	v7: Reference to political political supports or critics v1: Mention of basic assumptions/theoretical background
What are the putative reasons for the problem?	v7: Basic facts about the problem
What are the most important reasons?	v7: Basic facts about the problem
What are the indicators? (reliable, specific, realistic)	V12: Description of the sources of information Description of the instruments or way of measurement Description of the quality of the sources
Which indicators can be measured by the person that implements the project?	V12: Description of the sources of information
Who can take part in the project?	v14: Description of the sample/the unit of analysis
How many people out of the population of all the people that need this intervention can take part?	v14: Description of the formal procedure of drawing the sample Information about changes in the sample v11: description of the process of data collection
Is the attainment of the target group measured and documented?	v14: information about changes in the sample v15: dealing with missing data
How often should the indicators be measured? Which subjects should be included?	v17: at least two times of measurement (Pre-Post) Control group
Is there a journal of the project?	V20: Assuring that all information is as free from error as is possible and kept secure
Does the report meet all relevant expectations?	This is described by the whole mapping sentence.
Have all involved persons been informed and integrated?	> *utility*
What are the strengths of the applicant (the one who wants to get money for his project)?	> *feasibility*
Which resources are already available; and which are required in addition?	> *feasibility*
Are there enough resources for the marketing?	> *feasibility*
Who heads the project?	> *feasibility*
Who has to do what and when so that the project goes according to schedule?	> *feasibility*

4.4 Future directions

4.4.1 Elaboration of the model

According to Canter (1985), a mapping sentence can be the start as well as the conclusion of a research project. One way of thinking about a piece of facet research is to consider it as a process of refinement, elaboration, and validation of a mapping sentence.
The model in the present study was developed following the Standards of the Joint Committee (1994), Rossi et al. (2002), and the standards formulated by the Campbell Collaboration (Farrington, 2003). The structure of the variables was taken over from the literature to the mapping sentence. In particular, this means that the standard-facet contains 30 elements/variables that might be described more clearly in facet-theoretic terms by moving away from the structure inherent in the literature.

Based on the knowledge gained from the facet-theoretic analyses, the initial mapping sentence can be elaborated, so that the structure of the evaluation criteria will become more obvious. The additional interpretation of the MDS space (see Figure 14) demonstrates that each variable under study can be described as a combination of two different aspects:

1. content aspects, namely, whether the information is geared either to needs of practitioners or to scientific standards
2. criteria of quality, namely, methodological accuracy contrasted to explicitness/ comprehensiveness

Following this idea, these two aspects could be separated from each other, in order to simplify the model. In this way, the large standard-facet with 30 elements could be split into two smaller facets. These new facets could then be used one by one, and their role in the process of assessment of quality could be investigated.

At the present stage of research, the author can report only initial ideas for a collection of elements for each of the two possible facets:
A standardised scientific evaluation report, as can be found in international journals, might be an interesting starting point for a collection of elements for a facet of content aspects. This is because a standardised scientific report should contain all relevant content aspects geared to scientific standards. A first attempt to visualise these aspects in a hierarchical order is presented in Figure 17.
Additionally, this facet should contain aspects that are interesting for practitioners. The present model already covers two such variables. These variables deal with the political, social and economic context. For additional ideas, the researcher could refer to guidelines for practitioners (e.g., Hupfeld, 2004).

Figure 17:
Structure of a scientific report

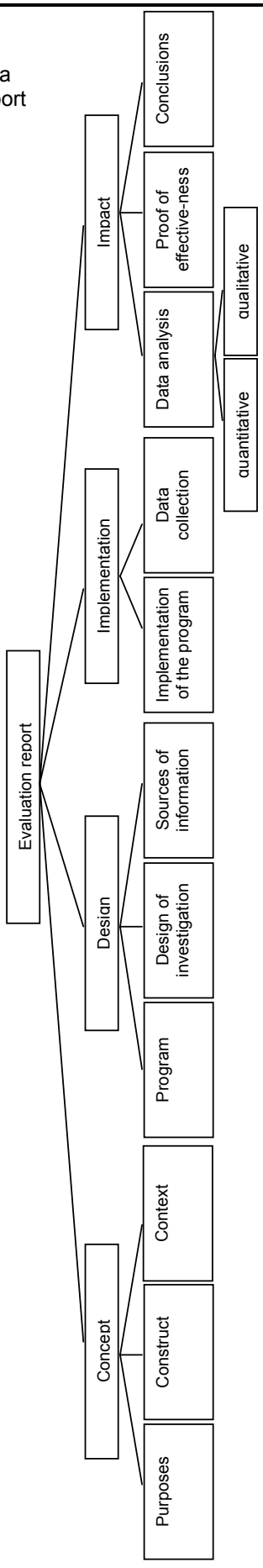
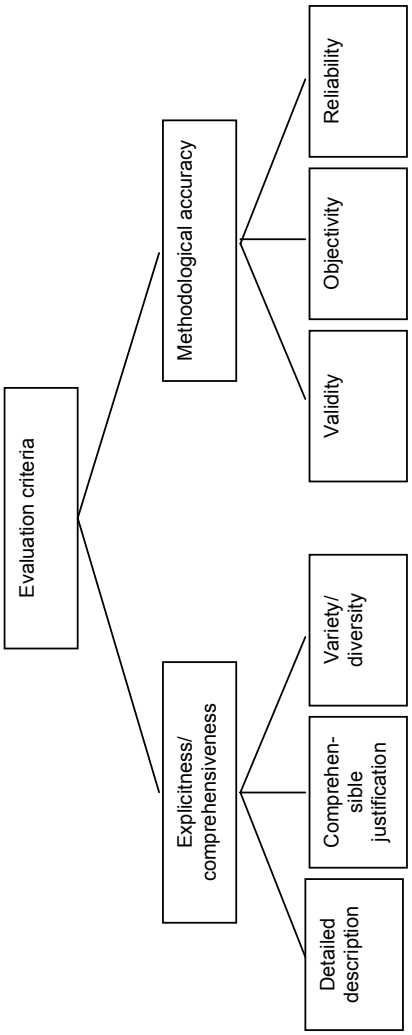


Figure 18:
Structure of evaluation criteria



For the second facet, quality criteria can be split into methodological accuracy and explicitness/comprehensiveness. These can be specified once again, as can be seen in Figure 18.
In future research, these initial ideas have to be elaborated, in order to create a comprehensive and refined set of elements for each facets. The presentation in terms of facets and the specification of their interrelationship in a mapping sentence might be interesting additional steps.

4.4.2 Generalisation

Apart from elaborating the mapping sentence, future research should also move on towards generalisability of the findings. Basically, there arise two main ways to generalise:
Firstly, the model covers only the aspect of accuracy, which seems to be the core aspect in treating evaluation reports. Nonetheless, the comparison with Hupfeld shows that utility and feasibility as other major parts in the *Standards* by the *Joint Committee* are of essential practical relevance. Therefore, it is desirable to introduce utility and feasibility (or even the fourth main aspect in the *Standards*: propriety).
Secondly, as mentioned in the Introduction, the mapping sentence that was developed in this study should be applicable to evaluations in different content areas. Yet, until now it has only been applied to a small sample of crime prevention reports as one specific sample.
This first application of facet theory to program evaluation standards has shown that it is a fruitful approach. But, future research is needed, to validate and to generalise the results and to elaborate practical implications.

References

Bilsky, W., & Cairns, D. (in press). Facettentheorie. In H. Holling & R. Schwarzer (Eds.), Enzyklopädie der Psychologie: Sec. B/IV/1. *Evaluation. Grundlagen und Methoden der Evaluationsforschung*. Göttingen: Hogrefe.

Borg, I. (1992). *Grundlagen und Ergebnisse der Facettentheorie.* Bern: Huber.

Borg, I. (1996). Facettentheorie. In E. Erdfelder, R. Mausfeld, T. Meisner, & G. Rudinger (Eds.), *Quantitative Methoden der Psychologie* (pp. 231-240). München: Urban & Schwarzenberg.

Borg, I. & Groenen, P. (1997). *Modern multidimensional scaling. Theory and application.* Berlin: Springer.

Borg, I. & Shye, S. (1995). *Facet Theory: Form and content.* Thousand Oaks: Sage.

Borg, I., & Staufenbiel, T. (1993). *Theorien und Methoden der Skalierung. Eine Einführung*. Bern: Hans Huber.

Bühl, A, & Zöfel, P. (2002). *Erweiterte Datenanalyse mit SPSS*. Wiesbaden: Westdeutscher Verlag.

Campbell, D. T., & Stanley, J. C. (1966). *Experimental and quasi-experimental designs for research.* Chicago: Rand McNally.

Canter, D. (Ed.).(1985). *Facet Theory.* New York: Springer.

Chen, H., & Rossi, P. H. (1980). The multi-goal, theory-driven approach to evaluation : a model linking basic and applied social science. *Social Forces, 59,* 106-122.

Chen, H., & Rossi, P. H. (1983). Evaluating with sense : the theory-driven approach. *Evaluation Review, 7*, 283-302.

Cook. T. D., & Campbell, D. T. (1979). *Quasi-experimentation: Design and analysis issues for field settings*. Chicago: Rand McNally.

Cook, T. D., & Matt, G. E. (1990). Theorien der Programmevaluation – Ein kurzer Abriß. In U. Koch & W. Wittmann (Eds.), *Evaluationsforschung* (pp. 15-38). Berlin: Springer.

Dancer. L. S. (1990). Introduction to facet theory and its applications. *Applied Psychology, 39,* 365-377.

Farrington, D. P. (2001). Editorial. Systematic reviews of criminological interventions. *Criminal Behavior and Mental Health, 11,* 127-130.

Farrington, D. P. (2003). Methodological Quality Standards for Evaluation Research. *ANNALS, AAPSS (American Academy of Political and Social Science), 587,* 49-68.

Guttman, L. (1959). Introduction to facet design and analysis. In *Proceedings of the Fifteenth International Congress of Psychology, Brussels, 1957* (pp. 130-132). Amsterdam: North Holland.

Guttman, L. (1968). A general nonmetric technique for finding the smallest coordinate space for a configuration of points. *Psychometrika, 33,* 469-506.

Guttman, R. & Greenbaum, C. W. (1998). Facet theory: Its development and current status. *European Psychologist, 3*, 13-36.

Home Office (2002). *Passport to evaluation.* Retrieved in October 2003 from http://www.crimereduction.gov.uk/learningzone/passport_to_evaluation.htm.

Howell, D.C. (2002). *Statistical Methods for Psychology* (5th ed.). Boston: Duxbury Press.

Hupfeld, J. (2004). *Kommunale Kriminalprävention: Ein Leitfaden zum Qualitätsmanagement bei der Planung, Durchführung und Evaluation kriminalpräventiver Projekte.* Düsseldorf: Landespräventionsrat Nordrhein-Westfalen.

Joint Committee on Standards for Educational Evaluation, Sanders, J. R. (Ed.). (1994). *The Program Evaluation Standards (2nd ed.).*Thousand Oaks: Sage.

Landeshauptstadt Düsseldorf (Eds.).(2001). *Düsseldorfer Gutachten: Empirisch gesicherte Erkenntnisse über kriminalpräventive Wirkungen.* Retrieved in October 2003 from http://www.duesseldorf.de/download/dg.pdf.

Levy, S. (1985). Lawful roles of facets in social theories. In D. Canter (Ed.), *Facet Theory* (pp. 59-96). New York: Springer.

Levy, S. (Ed.).(1994). *Louis Guttman on theory and methodology: Selected writings.* Aldershot: Dartmouth.

Levy, S. & Bar-On, E. C. (2003). A partly ordered typology of drug use in Israel. In S. Levy & D. Elizur (Eds.), *Facet Theory – Towards Cumulative Social Science* (pp.357-365). Ljubljana: University of Ljubljana.

Lipsey, M. W. & Wilson, D. B. (2001). *Practical meta-analysis*. Thousand Oaks: Sage.

Rosenthal, R. (1983). Meta-analysis: toward a more cumulative social science. In L. Bickman (Ed.), *Applied Social Psychology Annual 4* (pp. 65-94). London: Sage.

Rossi, P. H., Freeman, H. E., & Lipsey, M. W. (2002). *Evaluation.* Thousand Oaks: Sage.

Shadish, W. R, Cook, T. D., & Campbell, D. T. (2002). *Experimental and quasi-experimental designs for generalized causal inference.* Boston: Houghton-Mifflin.

Sherman, L. W., Gottferdson, D., MacKenzie, D., Eck, J., Reuter, P., & Bushway, S. (1997). *Preventing crime: What works, what doesn't, what's promising.* University of Maryland, department of Criminology and Criminal Justice. Retrieved in October 2003 from http://www.ncjrs.org/works/.

Shye, S. (1985). *Multiple Scaling. The theory and application of Partial Order Scalogram Analysis.* Amsterdam: North Holland.

Shye, S. & Elizur, D. (1994). *Introduction to Facet Theory: Content design and intrinsic data analysis in behavioral research.* Thousand Oaks: Sage.

Spence, I., & Ogilvie, J. C. (1973). A table of expected stress values for random rankings in nonmetric multidimensional scaling. *Multivariate Behavioral Research, 8,* 511-517.

Süß, H.-M. J. (2003). *Gutachterliche Stellungnahme zu den Berichten der extern evaluierten kriminalpräventiven Projekte.* Unpublished manuscript.

Van de Geer, J. P. (1993). *Multivariate analysis of categorical data: theory*. Newbury Park: Sage.

Weiss, C. H. (1973). Where politics and evaluation meet. *Evaluation, 1*, 37-45.

Weiss, C. H. (1975). Evaluation research in the political context. In E. L. Struening, & M. Guttenrag (Eds.), *Handbook of evaluation research,* Vol.1 (pp. 13-26). Beverly Hills: Sage.

Wilson, M. (1995). Structuring qualitative data: Multidimensional scalogram analysis. In G.M. Breakwell, S. Hammond, and C. Fife-Schaw (Eds.), *Research methods in psychology* (pp.259-273). London: SAGE Publications.

Wholey, J. S. (1979). *Evaluation: promise and performance*. Washington/DC: Urban Institute.

Wholey, J. S. (1983). *Evaluation and effective public management*. Boston: Little, Brown.

Zentrale Geschäftsstelle Polizeiliche Kriminalprävention der Länder und des Bundes (Eds.).(2003). *Handbuch Evaluation: Qualitätssicherung Polizeilicher Präventionsprojekte. Eine Arbeitshilfe für die Evaluation.* Retrieved in October 2003 from http://www.polizei.propk.de/mediathek/fachpublikationen/pdf/evaluation.pdf.

Zvulun, E. (1978). Multidimensional Scalogram Analysis: The Method and Its Applications. In S. Shye (Ed.). *Theory construction and data analysis in the behavioural sciences* (pp. 237-264). San Francisco: Jassey-Bass Publishers.

Appendix

Shlomit Levy: Suggestion for mapping sentences

Mapping sentence 1[a]

Respondent (x) assesses the extent of use by himself of				(1. legal) (2. illegal)	chemical substances	
Of the kind	(1. Psychoactive medications) (2. Stimulants) (3. Methadone) (4. Marijuana) (5. Heroin) (6. Cocaine) (7. L.S.D.) (8. Ecstasy) (9. Codeine-ephedrine) (10. G.H.D.) (11. Poppers (volatile substances))	with	(1. daily) (2. weekly) (3. monthly) (4. opportunistic) (5. seldom) (6. never) (7. unspecified)		frequency	
→	(high) (to) (low)	use of chemical substances (drugs).				

[a] personal communication between Shlomit Levy and David Cairns (03/29/04). The mapping sentences are suggestions for a facet-theoretic model for an analysis in:
Levy, S. & Bar-On, E. C. (2003). A partly ordered typology of drug use in Israel. In S. Levy & D. Elizur (Eds.), *Facet Theory – Towards Cumulative Social Science* (pp.357-365). Ljubljana: University of Ljubljana.

Mapping sentence 2[a]

	A							
	(1. instrumental)							
The	(2. cognitive)	attitude of respondent (x) concerning						
	(3. affective)							
	B							
	(1. physical)		C			D		
	(2. psychological)		(1. damage)			(1. himself)		
the	(3. social)		(2. advantage)		of the use by	(2. others)		
	(…)		(3. unspecified)			(3. unspecified)		
	(unspecified)							
						F		
						(1. Psychoactive medications)		
						(2. Stimulants)		
						(3. Methadone)		
						(4. Marijuana)		
	E					(5. Heroin)		
of	(1. legal)	chemical substances/drugs of the kind				(6. Cocaine)		
	(2. illegal)					(7. L.S.D.)		
						(8. Ecstasy)		
						(9. Codeine-ephedrine)		
						(10. G.H.D.)		
						(11. Poppers (volatile substances))		
	G							
	(1. daily)							
	(2. weekly)							
	(3. monthly)				(positive)			
with	(4. opportunistic)		frequency	→	(to)	attitude towards		
	(5. seldom)				(negative)			
	(6. never)							
	(7. unspecified)							

drug abuse/chemical sustances.

Coding frame

General notes

- *The Program Evaluation Standards (1994)* published by *The Joint Committee on Standards for Educational Programs* served as a general framework for this coding frame.
- The grouping of the elements in this coding frame (A1-A11) directly follows the structure of the part *Accuracy Standards*.
- The 30 elements have been derived from the overview text and the guidelines provided for each standard.
- The information for clusters A1 (Program documentation) and A5 (Valid information) provided in *The Program Evaluation Standards (1994)* did not seem to be detailed enough to allow for coding. Therefore, the coding instructions for the elements in these clusters contain some additional information. For A1 (Program documentation), the coding frame refers to Rossi, Freeman and Lipsey (2002). For A5 (Valid information), it refers to guidelines provided by the Campbell Collaboration (Source: Farrington, David P.2003. Methodological Quality Standards for Evaluation Research. *Annals of the American Academy of Political and Social Science* 587:49-68.)
- In order to distinguish common statistical approaches from sophisticated ones, the coding frame refers to the following internationally accepted introductory statistics book:
  Howell, D.C. (2002). *Statistical Methods for Psychology* (5th ed.). Boston: Duxbury Press.

A1 Program documentation

The program being evaluated should be described and documented clearly and accurately.

1. Description of the unique features of the program

The following text passages are excerpts from Rossi (1999). The criteria that the coding frame refers to are highlighted.

The foundation on which every program rests is some conception of what must be done to bring about the intended social benefits, whether that conception is expressed in a detailed program plan and rationale or is only implicit in the program's structure and activities. That conception is what we have referred to as program theory.

The most important framework for assessing program theory builds on the results of needs assessment. Or, more generally, it is based on a thorough understanding of the social problem the program is intended to address and the service needs of the relevant target population. A program theory that does not embody a conceptualization of program activities and outcomes that relate in an appropriate and effective manner to the actual nature and circumstances of the social conditions at issue will yield an ineffective

program no matter how well implemented and administered.

A thorough job of articulating program theory should reveal for inspection the critical assumptions and expectations inherent in the program's design. The program's goals and objectives will be specified and the primary program components and functions will be identified.

Are the program goals and objectives well defined? The outcomes for which the program is to be accountable should be stated in sufficiently clear and concrete terms that it is possible to determine if they have been attained. One line of inquiry on this issue is to ask if there are observable implications of the goals and objectives such that meaningful measures and indicators of success could be defined.

basic assumptions/theoretical background	
1	• The social problem the program is intended to address is mentioned • *and* background (theories, connection to other programs)
0	• None or only one of the points above is mentioned
aims of the program	
1	• Aims of the program mentioned • *and* there are observable implications of the goals and objectives such that meaningful measures and indicators of success could be defined
0	• No aims mentioned or the aims are too broad to be measured
2. Description of the component parts	
content of the program (e.g., sessions, examples)	
1	• Broad description of the content (anything beyond a label)
0	• No description
3. Description of the implementation of the program	
discrepancies between design and implementation of the program	
1	• Any comparison between design and implementation *or* any further information about the implementation (something that was not mentioned in the design part)
0	• No comparison, just the design mentioned
4. Association of the components of the program with its effect	
1	• Reference to the program and/or its parts in the design or the data analysis (intended effect of the program); *or* in the interpretation, the reporting of outcomes *or* the discussion (real effects of the program)
0	• No reference to the program and/or its parts in the design or the data analysis (intended effect of the program); or in the interpretation, the reporting of outcomes or the discussion (real effects of the program)

A2 Context analysis	
The context in which the program exists should be examined in enough detail.	
5. location of the program	
geographic location	
1	• Geographic location mentioned (at least country and region)
0	• Geographic location not mentioned
setting	
1	• Setting mentioned • *and* described (at least some information about the setting that is related to the problem to be addressed or to the program > why was this setting chosen to test the program?)
0	• Setting not mentioned or just mentioned and not described
6. its timing	
When implemented	
1	• Date of implementation mentioned (at least: year)
0	• Date of implementation not mentioned
Time schedule	
1	• Duration of the whole program/project • *and* sequence of the different parts (if there are several parts) • *and* (if the program consists of several intervention units) number of units and time period between them (e.g. sessions)
0	• No description of time schedule
7. the political and social climate surrounding it	
political support or critics	
1	• Reference to political context at any stage of the report
0	• No reference to political context
basic facts about the problem	
1	• Description of the problem in general *or* in this setting (e.g. statistical data or narrative)
0	• No description of the problem
8. the staff	
1	• Important persons that are involved are mentioned (e.g., trainers, teachers, bus drivers) • *and* some statements on their qualification *and* motivation (not necessarily a test) > It is not important whether they are motivated and well-qualified. The important point is that some information/estimation is provided.
0	• staff is not mentioned or just mentioned but without further information about qualification and motivation
9. pertinent economic conditions	

1	• Reference to resources (time and/or money)
0	• No reference to resources (time and/or money)

A3 Described purposes and procedures
The purposes and procedures of the evaluation should be monitored and described in enough detail, so that they can be identified and assessed.
10. The evaluation purposes
objectives/hypotheses

1	• The objectives *or* the hypotheses of the evaluation are mentioned (Why was the evaluation necessary?)
0	• The objectives *or* the hypotheses of the evaluation are not mentioned

intended use of its results	
1	• intended use of its results mentioned
0	• intended use of its results not mentioned

11. Description of the procedures	
Description of data collection	
1	• Description of the process of data collection > instruments and drop-outs are not included (covered by other elements); to get a 1, a report must answer the following question: What was the general approach to data collection? (e.g., sending questionnaires to all participants, collecting data from police statistics)
0	• No description of the process of data collection
Description of data analysis	
1	• The methods of data analysis are mentioned. If they are very special (not mentioned in Howell (2002)), they must be explained.
0	• The methods of data analysis are not mentioned. Or they are very special, but not explained.

A4 Defensible information sources
The sources of information used in a program evaluation should be described in enough detail, so that the adequacy of the information can be assessed.
12. Description of the sources of information
Description of the instruments or way of measurement

1	• Type of instrument (e.g., questionnaire) mentioned • *and* variables that are measured by this instrument in this report
0	• none or just one of the points mentioned

Description of their quality	
1	• any measures of quality provided
0	• No measures of quality provided

13. Variety of sources	
1	• At least one variable was assessed by using different sources of information to allow for data loss
0	• No variable was assessed by using different sources of information to allow for data loss
14. Description of the sample	
Description of the formal procedure of drawing the sample	
1	• Description of the formal procedure of drawing the sample (e.g., random sampling vs. recruiting volunteers; exclusion of subjects)
0	• No description of the formal procedure of drawing the sample
Description of the sample/the unit of analysis	
1	• Sample size • *and* Characteristics: age *and* gender *and* relation to the problem / the setting (e.g., men who battered, children in a school class)
0	• none or just one of the points mentioned
Changes in the sample	
1	• Information about changes in the sample (new people joining the sample or dropouts); also possible: information that there was no change in the sample
0	• No information about changes in the sample
15. Dealing with missing data	
1	• Information about dealing with missing data (e.g., leaving out the subject who didn't fill in the whole questionnaire); also possible: information that there were no missing data
0	• No information about dealing with missing data

A5 Valid information	
The information gathering procedures should be chosen or developed and then implemented so that they will assure that the interpretation arrived at is valid for the intended use.	
16. Construct validity	
Adequacy of the operational definition	
1	• relation between theory and choice of variables/tests is explained > it is enough if the author says that a test he/she used was designed to assess one of his variables
0	• relation between theory and choice of variables/tests is not explained
Validity of the instruments	
1	• Report contains information on validity of the instrument that was used to assess (one of the) central construct(s) > coefficient provided
0	• Report contains no information on validity of the instrument that was used to assess (one of the) central construct(s)

Multiple sources of information (convergent validity) ➢ overlapping with A4 13 Variety of Sources ➢ difference: the different sources are used to increase validity (comparison between them is necessary)	
1	• At least one variable was assessed by using different sources of information to increase validity
0	• No variable was assessed by using different sources of information to increase validity
Assessment of unintended effects (most important in the context of crime: displacement)	
1	• Unintended effects were assessed (unintended = the program showed an effect; nevertheless the problem was not solved) > most important in the context of crime: displacement
0	• Unintended effects were not assessed
17. Internal validity	
Experimental manipulation (vs. correlational design)	
1	• Experimental manipulation
0	• No experimental manipulation
Control condition	
1	• There is a control group
0	• There is no control group
Random assignment	
1	• Random assignment
0	• No random assignment • *or* unclear
At least two times of measurement (Pre-Post)	
1	• At least two times of measurement (Pre-Post)
0	• Less than two times of measurement (Pre-Post)
Blind participants and blind observers	
1	• Blind participants • *and* If there are observers, they must be blind as well
0	• no blind participants or blind participants but no blind observers • *or* unclear
Other possible influences or mediators were assessed	
1	• At least one possible influence was assessed (e.g., social desirability) *or* one possible mediator
0	• No possible influence and no possible mediator were assessed
18. Statistical conclusion validity	
Appropriate statistical methods	
1	• As far as one can tell from the report: The data do not violate the underlying assumptions of the statistical test(s)

0	• As far as one can tell from the report: The data violate the underlying assumptions of the statistical test(s)
Calculation of effect sizes or equivalents	
1	• At least one effect size is calculated • *or* one equivalent (equivalent = means and standard deviations of experimental and control group in pre- and post-test)
0	• No effect size is calculated and no equivalent

A6 Reliable information	
The information gathering procedure should be chosen or developed and then implemented so that they will assure that the information obtained is sufficiently reliable for the intended use.	
19. Assessment of the reliability of the instruments	
1	• The reliability of the instrument measuring the central construct was assessed (e.g., inter-rater-reliability, reliability of tests, stability) > coefficient provided
0	• The reliability of the instrument measuring the central construct was not assessed

A7 Systematic information	
The information collected, processed, and reported in an evaluation, should be systematically reviewed and any errors found should be corrected.	
20. Assuring that all information is as free from error as is possible and kept secure	
1	• The report mentions any kind of attempts that were made to find and correct errors. • Examples mentioned in *The Program Evaluation Standards*: Ensuring that evaluation staff are adequately trained to carry out their roles and are sensitized to the kind of mistakes they are likely to make; systematically checking for errors in the collecting, processing, and reporting of information and results; using another person to verify data entry; monitoring outside agencies or individuals responsible for information collecting, scoring and categorization, and/or quantitative or qualitative analyses; maintaining control of original information and results so that their integrity can be protected; adopting and implementing standard procedures for storing and retrieving information; checking with stakeholders routinely to make certain information collected from them is represented accurately.
0	• The report does not mentions any attempts that were made to find and correct errors

A8 Analysis of quantitative information
Quantitative information in an evaluation should be appropriately and systematically analysed so that evaluation questions are effectively answered.

21. Use of initial exploratory (descriptive) analyses	
1	• This covers any kind of analyses that assess the nature or the quality of the data, any descriptive statistics (e.g. distributions, scatterplots)
0	• No initial exploratory analyses mentioned
22. Use of more sophisticated and complex analyses	
This covers any kind of analyses that go beyond the initial exploratory analyses. > inferential	
Mentioning and describing the procedure used	
1	• A “sophisticated and more complex” procedure was used (e.g., ANOVA, regression, t-test, LISREL) • *and* If it is not described in an introductory statistics book (Howell (2002)), there must be a broad description (general idea, why used here, prerequisites); exception: factor analysis
0	• No “sophisticated and more complex” procedure was used
Giving appropriate statistics	
1	• The common statistics for this analysis are given > common means that they would be in the standard SPSS output for this analysis (e.g., F, df, mean standard deviation) • *and* If this analysis is not described in an introductory statistics book (Howell (2002)), there must be a short explanation of the scores and their meaning; exception: factor analysis
0	• The common statistics for this analysis are not given; or they are not common and not explained
23. Visual displays	
1	• The report contains at least one visual display of quantitative information that the author refers to in the text (graphs, tables etc.)
0	• The report contains no visual display; or the author does not refer to it in the text.

A9 Analysis of qualitative information
Qualitative information in an evaluation should be appropriately and systematically analysed so that evaluation questions are effectively answered. *Definition of “qualitative” (according to THE JOINT COMMITTEE ON STANDARDS FOR EDUCATIONAL EVALUATION): Qualitative information consists of descriptions and interpretations that are in narrative rather than numerical form.* *e.g, interviews, observations, hearings, documents* *The advantage of including qualitative information (according to THE JOINT COMMITTEE ON STANDARDS FOR EDUCATIONAL EVALUATION): Qualitative analysis may give depth and perspective to the data that quantitative analysis alone is not able to provide.* *The essence of this standard is both to enrich understanding of the phenomena under study and to avoid faulty conclusions.*

24. set of categories / approach towards structuring the qualitative data *> structuring the data and making the analysis comprehensible*	
1	• The report contains qualitative information; and a set of categories has been derived.
0	• No set of categories
25. Assessment of the quality of categories / structure *> quality of the qualitative analysis*	
1	• The report contains qualitative information; a set of categories has been derived; and their quality has been assessed in some way (e.g., validity, reliability)
0	• No assessment of quality of the categories
26. meaningfulness of conclusions and recommendations *> enrich understanding of the phenomena under study*	
1	• The report contains qualitative information; and the meaningfulness of conclusions and recommendations has been demonstrated with reference to the qualitative information = reference to the qualitative information in the conclusions or recommendations
0	• No reference to the qualitative information in the conclusions or recommendations

A10 Justified Conclusions	
The conclusions reached in an evaluation should be explicitly justified, so that the stakeholders can assess them.	
27. Adequate interpretation of statistics	
1	• The values of the important statistics (at least the main result for each statistical procedure > e.g. significance level) included in the report are interpreted (to interpret = any meaning attached to the value) • *and* This interpretation is adequate in that the meaning attached to a value is appropriate with respect to its value and statistical background (e.g., significance because of large sample size)
0	• No interpretation of statistics or an interpretation that is inadequate
28. Relation of conclusions to statistical results	
1	• The report relates the main conclusions to the statistical results • *and* The conclusions do not go beyond the (statistical) results, unless they are explicitly characterised as going beyond the (statistical) results > specs
0	• The report does not relate the main conclusions to the statistical results
29. Possible alternative explanations for results	
1	• Possible (conflicting) alternative explanations for the results mentioned
0	• No possible alternative explanations for the results mentioned
A11 Impartial reporting	

Reporting procedures should guard against distortion caused by personal feelings and biases of any party to the evaluation, so that evaluation reports fairly reflect the evaluation findings. *Possible conflicting explanations for the results and adequate interpretations are already covered by A10 (Justified conclusions). Furthermore, a report would be biased if not everything that is contrary to the program was reported. As you cannot tell from a report which facts were not included, this coding plan focuses on the language.*	
30. Neutral and objective style of reporting at any stage of the report	
1	• Neutral language (no words expressing personal feelings); if there are any evaluative words in the text, they are included in the introduction, the interpretation, or the discussion
0	• Language is not neutral

Coding sheet

Number of Report:

Title:
Author:
Year of Publication:
Country:

Date of Coding:

	Range of code	Code
A1 Program documentation		
1. Description of the unique features of the program		
Basic assumptions/theoretical background	0,1	
Aims of the program	0,1	
Sum	0,1,2	
2. Description of the component parts		
Content of the program	0,1	
3. description of the implementation of the program		
Discrepancies between design and implementation of the program	0,1	
4. Association of the components of the program with its effect		
	0,1	
A2 Context Analysis		
5. location of the program		
Geographic location	0,1	
Setting	0,1	
Sum	0,1,2	
6. its timing		
When implemented	0,1	
Time schedule	0,1	
Sum	0,1,2	
7. The political and social climate surrounding it		

Political support or critics	0,1	
Basic facts about the problem	0,1	
Sum	0,1,2	
8. The staff		
	0,1	
9. Pertinent economic conditions		
	0,1	
A3 Described purposes and procedures		
10. The evaluation purposes		
Objectives/hypotheses	0,1	
Intended use of its results	0,1	
Sum	0,1,2	
11. Description of the procedures		
Description of data collection	0,1	
Description of data analysis	0,1	
Sum	0,1,2	
A4 Defensible information sources		
12. Description of the sources of information		
Description of the instrument or way of measurement	0,1	
Description of their quality	0,1	
Sum	0,1,2	
13. Variety of sources		
	0,1	
14. Description of the sample		
Description of the formal procedure of drawing the sample	0,1	
Description of the sample/the unit of analysis	0,1	
Changes in the sample	0,1	
Sum	0,1,2,3	
15. Dealing with missing data		

	0,1	
A5 Valid information		
16. Construct validity		
Adequacy of the operational definition	0,1	
Validity of the instruments	0,1	
Multiple sources of information	0,1	
Assessment of unintended effects	0,1	
Sum	0,1,2,3,4	
17. Internal validity		
Experimental manipulation	0,1	
Control condition	0,1	
Random assignment	0,1	
At least two times of measurement (Pre-Post)	0,1	
Blind participants and blind observers	0,1	
Other possible influences or mediators were assessed	0,1	
Sum	0,1,2,3,4,5,6	
18. Statistical conclusion validity		
Appropriate statistical methods	0,1	
Calculation of effect sizes or equivalent	0,1	
Sum	0,1,2	
A6 Reliable information		
19. Assessment of the reliability of the instruments		
	0,1	
A7 Systematic information		
20. Assuring that all information is as free from error as is possible and kept secure		
	0,1	
A8 Analysis of quantitative information		
21. Use of initial exploratory (descriptive) analyses		
	0,1	

22. Use of more sophisticated and complex analyses		
Mentioning and describing the procedure	0,1	
Giving appropriate statistics	0,1	
Sum	0,1,2	
23. Visual displays		
	0,1	
A9 Analysis of qualitative information		
24. set of categories/ approach towards structuring the qualitative data		
	0,1	
25. Assessment of the quality of categories / structure		
	0,1	
26. meaningfulness of conclusions and recommendations		
	0,1	
A10 Justified conclusions		
27. Adequate interpretation of statistics		
	0,1	
28. Relation of conclusions to statistical results		
	0,1	
29. Possible alternative explanations for results		
	0,1	
A11 Impartial reporting		
30. Neutral and objective style of reporting at any stage of the report		
	0 (not neutral), 1 (neutral)	

Comments:

Data set – ordinally scaled data

number	v1	v2	v3	v4	v5	v6	v7	v8	v9	v10	v11	v12	v13	v14	v15	v16	v17	v18	v19	v20	v21	v22	v23	v24	v25	v26	v27	v28	v29	v30
1	2	1	1	1	2	1	2	0	1	1	2	2	0	2	0	2	4	1	1	1	1	2	1	0	0	0	1	1	0	1
2	2	1	1	1	2	2	1	0	0	1	2	2	0	3	0	2	5	1	1	1	1	1	1	1	1	1	1	1	0	1
3	2	1	1	1	1	1	2	0	1	1	2	1	0	2	0	4	4	1	0	1	1	1	1	0	0	0	1	1	1	1
4	2	1	1	1	1	1	2	0	0	1	1	0	0	2	0	0	2	1	0	0	1	0	1	1	0	1	0	0	0	0
5	2	1	0	1	2	1	2	0	1	1	2	0	0	1	0	3	4	2	0	0	1	1	1	0	0	0	1	1	1	1
6	2	1	0	0	1	1	2	0	1	1	2	0	0	2	0	1	1	1	0	0	1	0	0	0	0	0	1	1	0	1
7	2	1	1	1	1	2	2	0	1	2	2	2	0	3	0	2	3	1	0	0	1	2	1	1	0	1	1	1	0	1
8	2	1	1	1	2	2	2	1	1	2	2	1	0	3	0	2	3	1	0	0	1	1	1	0	0	1	1	1	0	1
9	2	1	1	1	2	2	1	0	1	1	2	1	0	3	1	3	5	1	0	1	1	2	1	1	0	1	1	1	1	1
10	2	1	1	1	2	0	1	0	0	1	2	2	0	2	0	2	4	1	1	1	1	2	1	1	1	1	1	1	1	1
11	2	1	1	1	2	2	2	0	0	1	2	1	0	2	0	1	5	1	0	1	1	1	1	0	0	0	1	1	0	1
12	2	1	1	1	2	2	0	0	0	1	2	1	0	0	0	2	4	1	0	0	1	2	1	0	0	0	1	1	0	1
13	2	1	1	1	1	1	0	0	0	1	2	1	0	2	0	1	3	1	0	0	1	2	1	0	0	0	1	1	1	1
14	2	1	1	1	0	1	2	0	0	2	2	1	0	1	0	2	4	1	0	0	0	2	0	0	0	1	1	1	0	1
15	1	0	0	0	2	0	1	0	0	1	2	1	0	0	0	1	0	1	0	0	1	0	1	0	0	0	1	0	0	0
16	1	1	1	1	1	1	0	0	0	1	2	1	0	2	0	1	1	1	0	1	1	0	1	0	0	1	1	0	0	0
17	2	1	1	1	1	2	0	0	1	2	2	2	0	3	0	1	3	2	1	0	1	2	1	0	0	1	1	1	1	0
18	2	1	1	1	1	2	2	1	1	2	1	2	0	3	0	1	2	1	0	1	1	0	1	0	0	1	1	1	0	1
19	2	1	1	1	2	2	2	0	0	1	2	2	0	3	0	2	4	1	1	1	1	1	1	0	0	0	1	1	0	1
20	1	1	0	0	1	2	2	0	1	1	2	0	0	2	0	0	0	1	0	1	1	0	1	0	0	0	1	0	0	0
21	1	0	0	1	1	1	1	0	0	1	1	1	0	0	0	2	2	1	0	0	1	0	1	0	0	0	0	0	0	0
22	2	1	0	1	2	1	1	0	0	0	2	1	0	1	0	1	1	1	0	1	1	0	0	1	0	0	1	0	0	1
23	2	1	1	1	2	1	2	0	1	2	2	1	1	2	0	2	1	1	0	1	1	2	1	1	0	1	1	1	0	0
24	1	0	0	0	1	1	1	0	0	1	2	1	0	1	0	2	3	1	0	0	1	2	1	1	0	1	1	1	0	1
25	1	1	0	0	1	1	0	0	0	0	2	1	1	3	0	2	4	1	0	0	1	2	1	1	0	1	1	1	0	1

Rules for dichotomising

The following variables have more than one indicator

- V1 Description of the unique features of the program
- V5 Location of the program
- V6 Its timing
- V7 The political and social climate surrounding it
- V10 The evaluation purposes
- V11 Description of the procedures
- V12 Description of the sources of information
- V14 Description of the sample
- V16 Construct validity
- V17 Internal validity

- V18 Statistical conclusion validity
- V22 Use of more sophisticated and complex analyses

In general, a one was assigned to a report, if all indicators had been coded as one. There were two exceptions:
Because of the large number of indicators for v17 (internal validity), a one was assigned to a report, if all but two indicators had been coded as one.
The indicators for v16 (construct validity) were thought to be various possibilities to ensure construct validity rather than essential features of a good evaluation report. Therefore, a one was assigned to a report that had been coded one on at least two indicators.

Data set – dichotomous data

number	v1	v2	v3	v4	v5	v6	v7	v8	v9	v10	v11	v12	v13	v14	v15	v16	v17	v18	v19	v20	v21	v22	v23	v24	v25	v26	v27	v28	v29	v30
1	1	1	1	1	1	0	1	0	1	0	1	1	0	0	0	1	1	0	1	1	1	1	1	0	0	0	1	1	0	1
2	1	1	1	1	1	1	0	0	0	0	1	1	0	1	0	1	1	0	1	1	1	0	1	1	1	1	1	1	0	1
3	1	1	1	1	0	0	1	0	1	0	1	0	0	0	0	1	1	0	0	1	1	0	1	0	0	0	1	1	1	1
4	1	1	1	1	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	1	1	0	1	0	0	0	0
5	1	1	0	1	1	0	1	0	1	0	1	0	0	0	0	1	1	1	0	0	1	0	1	0	0	0	1	1	1	1
6	1	1	0	0	0	0	1	0	1	0	1	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	1	1	0	1
7	1	1	1	1	0	1	1	0	1	1	1	1	0	1	0	1	0	0	0	0	1	1	1	1	0	1	1	1	0	1
8	1	1	1	1	1	1	1	1	1	1	1	0	0	1	0	1	0	0	0	0	1	0	1	0	0	1	1	1	0	1
9	1	1	1	1	1	1	0	0	1	0	1	0	0	1	1	1	1	0	0	1	1	1	1	1	0	1	1	1	1	1
10	1	1	1	1	1	0	0	0	0	0	1	1	0	0	0	1	1	0	1	1	1	1	1	1	1	1	1	1	1	1
11	1	1	1	1	1	1	1	0	0	0	1	0	0	0	0	0	1	0	0	1	1	0	1	0	0	0	1	1	0	1
12	1	1	1	1	1	1	0	0	0	0	1	0	0	0	0	1	1	0	0	0	1	1	1	0	0	0	1	1	0	1
13	1	1	1	1	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	1	1	1	0	0	0	1	1	1	1
14	1	1	1	1	0	0	1	0	0	1	1	0	0	0	0	1	1	0	0	0	0	1	0	0	0	1	1	1	0	1
15	0	0	0	0	1	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	1	0	1	0	0	0	1	0	0	0
16	0	1	1	1	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	1	1	0	1	0	0	1	1	0	0	0
17	1	1	1	1	0	1	0	0	1	1	1	1	0	1	0	0	0	1	1	0	1	1	1	0	0	1	1	1	1	0
18	1	1	1	1	0	1	1	1	1	1	0	1	0	1	0	0	0	0	0	1	1	0	1	0	0	1	1	1	0	1
19	1	1	1	1	1	1	1	0	0	0	1	1	0	1	0	1	1	0	1	1	1	0	1	0	0	0	1	1	0	1
20	0	1	0	0	0	1	1	0	1	0	1	0	0	0	0	0	0	0	0	1	1	0	1	0	0	0	1	0	0	0
21	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	1	0	1	0	0	0	0	0	0	0
22	1	1	0	1	1	0	0	0	0	0	1	0	0	0	0	0	0	0	0	1	1	0	0	1	0	0	1	0	0	1
23	1	1	1	1	1	0	1	0	1	1	1	0	1	0	0	1	0	0	0	1	1	1	1	1	0	1	1	1	0	0
24	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	1	0	0	0	0	1	1	1	1	0	1	1	1	0	1
25	0	1	0	0	0	0	0	0	0	0	1	0	1	1	0	1	1	0	0	0	1	1	1	1	0	1	1	1	0	1

List of evaluation reports

1.
Botvin, G. J. (1997). School-based drug abuse prevention with inner-city minority youth. *Journal of child and adolescent substance abuse, 6(1),* 5-20.

2.
Grossman, D. C., Neckerman, H. J., Koepsell, T. D., Liu, P., Asher, K. N., Beland, K. et al. (1997). Effectiveness of a violence prevention curriculum among children in elementary school. *Journal of the American Medical Association*, *277(20),* 1605-1611.

3.
Olweus, D. (1991). Bully-victim problems among schoolchildren: basic facts and effects of a school based intervention program. In D. J. Pepler (Ed.), *The development and treatment of childhood aggression* (pp. 411-448). Hillsdale, NJ: Erlbaum.

4.
Hanewinkel, R. (1999). Prävention von Gewalt an Schulen. In B. Röhrle & G. Sommer (Eds.), *Prävention und Gesandheitsförderung* (pp. 135-159). Tübingen: dgtv-Verlag.

5.
Olweus, D. (1996). Bullying at school: knowledge base and an effective intervention program. *Annals of the New York Axademy of Sciences, 794*, 265-276.

6.
Frances, R. (1995). An overview of community-based intervention programmes for men who are violent or abusive in the home. In R. E. Dobash, R. P. Dobash, & L. Noaks (Eds.), *Gender and crime* (pp. 390-409). Cardiff: University of Wales Press.

7.
Farley, D., & Magill, J. (1988). An Evaluation of a group program for men who batter. In G. S. Getzel (Ed.), *Violence. Prevention and treatment in groups* (pp. 53-65). New York/London: Haworth Press.

8.
Van Andel, H. (1989). Crime prevention that works: The care of public transport in the Netherlands. *British Journal of Criminology*, *29(1),* 47-56.

9.
Painter, K., & Farrington, D. P. (1997). The crime reducing effect of improved street lighting: the Dudley project. In R. V. Clarke (Ed.), *Situational crime prevention: successful case studies*. Guilderland/New York: Harrow and Heston.

10.
Breckheimer, S. E., & Nelson, R. O. (1976). Group methods for reducing racial prejudice and discrimination. *Psychological Reports, 39*, 1259-1268.

11.
Bennett, T. (1991). The effectiveness of a police-initiated fear-reducing strategy. *The British Journal of Criminology, 31(1),* 1-14.

12.
Stotland, E., Katz, D., & Patchen, M. (1959). The reduction of prejudice through the arousal of self-insight. *Journal of Personality, 27*, 507-531.

13.
Weiner, M. J., & Wright, F. E. (1973). Effects of undergoing arbitrary discrimination upon subsequent attitudes toward a minority group. *Journal of Applied Social Psychology, 3(1),* 94-102.

14.
Katz, J.H. (1977). White awareness: the frontier of racism awareness training. *The personnel and guidance journal, 55,* 485-489.

15.
Dollase, R. (2001). Die multikulturelle Schulklasse- oder: Wann ist der Ausländeranteil zu hoch? *Zeitschrift für politische Psychologie, 9,* 113-126.

16.
Der Landrat als Kreispolizeibehörde Gütersloh (2000). *Auswertung der* Befragung *zum Projekt Notruf-Handys für Senioren*. Unpublished manuscript.

17.
Molske, A. (2002). *Gewaltdeeskalationstraining für Lehrer/-innen*. Unpublished manuscript.

18.
Kober, M. (n.d.). *Die Präventionsplakette: Prozessevaluation eines Pilotprojekts der Kreisbehörde Gütersloh*. Unpublished manuscript.

19.
Voß, H.-G. W. (n.d.). *Professionalität in der polizeilichen Arbeit mit Opfern und Zeugen*. Unpublished manuscript.

20.
Hallmann, S., & Schmeitz, T. (1996). Befragung zu den Anti-Drogen-Diskotheken in Nordrhein-Westfalen. In D. Dölling (Ed.), *Drogenprävention und Polizei* (pp. 244-317). Wiesbaden: Bundeskriminalamt.

21.
Bathsteen, M., & Legge, I. (2001). Intendierte und nicht intendierte Folgen des Hamburger Substitutionsprogramms. *Monatsschrift für Kriminologie und Strafrechtsreform*, *84(1),* 1-9.

22.
Toprak, A. (2000). Kulturell bedingte Konflikte? – Anti-Aggressions-Kurse für männliche Jugendliche aus der Türkei. In E. Gropper & H.-M. Zimmermann (Eds.), *Raus aus Gewaltkreisläufen! Präventions- und Interventionskonzepte* (pp. 185-193). Aktion Jugendschutz (AJS), Jahrestagungsband. Landesarbeitsstelle Baden-Württemberg. Stuttgart.

23.
Bundesministerium für Frauen und Jugend (BMFJ) (1994). Sicherheitsbeitrag spezieller nächtlicher Beförderungsangebote (Disko-Busse). *Materialien zur Frauenpolitik*, 42. Bonn: Forschungsbericht des Planungshauses Südstadt AG Köln im Auftrag des BMFJ.

24.
Kraus, L., & Rolinski, K. (1992). Rückfall nach Sozialen Training auf der Grundlage offiziell registrierter Delinquenz. *Monatsschrift für Kriminologie und Strafrechtsreform*, *75(1),* 32-46.

25.
McCord, J. (1978). A thirty-year follow-up of treatment effects. *American Psychologist, 32,* 284-289.

Description of reports

Table D1: Description of reports

number	country	problem	approach	location
1	USA	Drugs	Offender	School
2	USA	Aggressiveness/violence	Offender	School
3	Norway	Aggressiveness/violence	Offender/victim	School
4	Germany	Aggressiveness/violence	Offender/victim	School
5	Norway	Aggressiveness/violence	Offender/victim	School
6	Australia	Aggressiveness/violence	Offender	Family/home
7	Canada	Aggressiveness/violence	Offender	Clinic
8	Netherlands	Vandalism/burglary	Offender/opportunity	Public transport
9	Great Britain	Crime in general	Opportunity	Neighbourhood
10	USA	Prejudice/discrmination	Offender	School
11	Great Britain	Fear of crime	Victim/witness	Neighbourhood
12	USA	Prejudice/discrimination	Offender	University
13	USA	Prejudice/discrimination	Offender	School
14	USA	Prejudice/discrimination	Offender	University
15	Germany	Prejudice/discrimination	Offender/opportunity/victim	School
16	Germany	Fear of crime	Victim/witness	Family/home
17	Germany	Aggressiveness/violence	Offender/opportunity/victim	School
18	Germany	Vandalism/burglary	Opportunity/victim	Family/home
19	Germany	Treatment of victims and witnesses	Victim/witness	Police
20	Germany	Drugs	Offender	School
21	Germany	Drugs	Offender	Aid organisations
22	Germany	Aggressiveness/violence	Offender	Clinic
23	Germany	Crime in general	Opportunity/victim	Public transport
24	Germany	Crime in general	Offender	Clinic
25	USA	Crime in general	Offender	Combination of locations

Berichte aus dem Psychologischen Institut IV

Aus der Arbeitseinheit "Differentielle Psychologie und Persönlichkeitspsychologie" sind bisher erschienen:

1/1994 WENTURA, D.: Gibt es ein "affektives Priming" im semantischen Gedächtnis?

2/1995 BILSKY, W.: Die Bedeutung von Furcht vor Kriminalität in Ost und West (unter diesem Titel in Monatsschrift für Kriminologie und Strafrechtsreform, 1996, 79, 357-372).

3/1996 BILSKY, W.: Ethnizität, Konflikt und Recht. Probleme von Assessment und Begutachtung in Strafverfahren mit Beteiligten ausländischer Herkunft. Antrag auf Sachbeihilfe bei der Volkswagenstiftung.

4/1996 BILSKY, W., BORG, I. & WETZELS, P.: La Exploración de Tácticas para la Resolución de Conflictos en Relaciones Íntimas: Reanálisis de un Instrumento de Investigación.

5/1997 BILSKY, W. & HOSSER, D.: Soziale Unterstützung und Einsamkeit: Zur Beziehung zweiter verwandter Konstrukte.

6/1997 BILSKY, W.: Vergleichende Strukturanalysen von Motiven und Werten.

7/1997 BILSKY, W.: Miedo al Delito, Victimizacion criminal, y la Relacion Miedo-Victimizacion: Algunos Problemas conceptuales y metodologicos.

8/1997 WENTURA, D.: The "meddling-in" of affective imformation: Evidence for negative priming and implicit judgement tendencies in the affective priming paradigm.

9/1997 BILSKY, W.: Strukturelle Beziehungen zwischen Motiven und Werten: Weitere Hinweise auf die Tragfähigkeit eines integrativen Modells.

10/1997 BILSKY, W.: Ethnizität, Konflikt und Recht. Bericht über ein von der Volkswagenstiftung im Schwerpunkt "Recht und Verhalten" gefördertes interdisziplinäres Symposium in der Werner-Reimers-Stiftung, Bad Homburg, vom 6.-8. Februar 1997.

11/1998 BILSKY, W.: Values and Motives. Paper presented at the International Research Workshop „Values: Psychological Structure, Behavioral Outcomes, and Inter-Generational Transmission“. Maale-Hachamisha, Israel, January 12-16th, 1998.

12/1998 BILSKY, W. & PETERS, M.: Estructura de los valores y la religiosidad. Una investigación comparada realizada en México.

13/1998 WENTURA, D.: Die Veränderung kognitiver Strukturen: Mikroprozessuale Aspekte der Bewältigung.

14/1998 WENTURA, D. & GREVE, W.: Adaptation und Stabilisierung selbstbezogener Kognitionen. Antrag auf Gewährung einer Sachbeihilfe an die Deutsche Forschungsgemeinschaft im Schwerpunktprogramm „Informationsverarbeitung im sozialen Kontext“.

15/1998 WENTURA, D. & NÜSING, J.: Situationsmodelle in der Textverarbeitung: Evidenz für die automatische Aktivierung emotional-entlastender Informationen.

16/1999 WENTURA, D.: Putting pieces together - or: Is there any relationship between „affective priming“ sensu Fazio et al. and „affective priming“ sensu Murphy and Zajonc.

17/1999 BILSKY, W. & JEHN, K. A.: Reconsiderations of value structures based on cross-cultural research: implications for organizational culture and conflict. Paper presented at the Twelfth Conference of the International Association for Conflict Management, June 20 - June 23, 1999, San Sebastián-Donostia, Spain.

18/1999 BILSKY, W. & RAHIM, M. A.: Mapping conflict styles – a facet approach. Paper presented at the Twelfth Conference of the International Association for Conflict Management June 20 – June 23, 1999, San Sebastián-Donostia, Spain.

19/1999 BILSKY, W.: Common structures of motives and values: towards a taxonomic integration of two psychological constructs.

20/1999 WENTURA, D., HOLLE, K. & KOMOGOWSKI, D.: Age stereotypes in younger and older woman; Analyses of accommodative shifts with a sentence-priming task.

21/2000 BILSKY, W. & WÜLKER, A.: Konfliktstile: Adaptation und Erprobung des ‚Rahim Organizational Conflict Inventory‘ (ROCI-II).

22/2000 BILSKY, W. & KOCH, M.: On the content and structure of values: Universals or methodological artefacts?

23/2002 BROCKE, M, GÖLDENITZ, C., HOLLING, H. & BILSKY, W.: Case characteristics and severity of punishment: Conjoint analytic investigations. Paper presented at the 12th European Conference on Psychology and Law, September 14 - September 17, 2002, Leuven, Belgium.

24/2002 BILSKY, W.: Fear of crime, personal safety and well-being: A common frame of reference. Paper presented at the 12th European Conference on Psychology and Law, September 14 - September 17, 2002, Leuven, Belgium.

25/2002 BILSKY, W.: La theoría de las facetas: Informaciones básicas y aplicaciones paradigmáticas.

26/2002 BILSKY, W. & BUBECK, M.: Value Structure at an Early Age.

27/2003 BILSKY, W., MÜLLER, J., VOSS, A. & VON GROOTE, E.: Measuring affect in crisis negotiation: An exploratory case study of hostage-taking.

28/2005 DÖRING, A.: A facet-theoretic approach.