

Finding Out What Works Best in the Measures Taken Against Right-Wing Extremism

Chantal Falk, Thomas Widmer and Cornelia Blaser

Department of Political Science, University of Zurich, Switzerland

falk@pwi.unizh.ch

September 13, 2004

Abstract

As part of a larger research project about right-wing extremism and its causes and countermeasures, we assess the quality of existing evaluation studies in Switzerland, Germany, the US and other countries in order to synthesize their results. This paper is presenting the results of the meta-evaluation in the field of measures against right-wing extremism in areas as education, social work, policing and others.

The meta-evaluation follows the professional evaluation standards set by the Swiss Evaluation Society (SEVAL Standards). The analysing of quality of selected evaluation studies in the field of measures against right-wing extremism allows us to assess the value of the results of these evaluations. The meta-evaluation consists systematic analysis of twelve evaluation reports. Based on these analyses, we will emphasize the conditions for success in evaluations. In addition this enables us to sketch a general picture of the current practice of evaluation of measures taken against right-wing extremism in selected countries. Moreover, this paper will present evidence about the usefulness of the SEVAL Standards in an applied setting.

Paper to be presented at the European Evaluation Society Sixth Conference, Berlin, September 30 – October 2, 2004.

Introduction

In recent years more extreme right-wing or xenophobic motivated (violent) crimes have been observed in Switzerland. Against this background, the former Head of the Federal Department of Justice and Police, Federal Council Mrs. Ruth Metzler, formed a study group “Right-wing Extremism”. The latter submitted a report on September 2000 and proposed to intensify the research activities in this area. Based on this recommendation, the Swiss Federal Government decided on June 2001 to give the Swiss National Science Foundation the task of expanding the National Research Programme “*Violence in Daily Life and Organized Crime*” (NRP 40) to include the topic “*Right-wing Extremism – Causes and Countermeasures*” by adding the NRP 40plus¹. The aim of NRP 40plus is to increase the knowledge of the causes, the profile, the extent and the consequences of right-wing extremist activities and attitudes in Switzerland. The results should establish a basis for future strategies for controlling and reducing right-wing extremism at municipal, cantonal and national levels. Furthermore, the programme intends to enhance links between research into right-wing extremism in Switzerland and scholars from other countries. The NRP 40plus started on August 2003 and runs over a period of three years. The programme centres around four research focuses, namely ‘right-wing extremist attitudes in both population and institutions’, ‘offenders and victims’, ‘social context’, and ‘evaluation of measures’.

The fourth research topic is expected to provide systematic evidence about the efficiency of measures against right-wing extremism. In addition, the projects within this research focus should provide findings in respect to suitable designs for evaluations of countermeasures. Since the beginning of the 1990s, in Switzerland as in other European countries, a variety of measures aimed to prevent and combat right-wing extremism have been implemented. In Switzerland, these measures consist mostly of municipal or cantonal initiatives in youth work, education and public campaigns, but as well repressive measures as the an anti-racism article in the Swiss Federal Penal Code (StGB Art 261bis). Furthermore the Swiss Federal Government is supporting inter alia educational initiatives or help lines. Measures of this kind are difficult to evaluate, not only due to the many different forms they take (and to the fact that as a social phenomenon, right-wing extremism displays a very complex interplay of social, economic, and political factors), but also because the goals of the measures themselves tend to be set in the long term.

¹ see www.nfp40plus.ch (in French and German only) for further information about this research programme.

Our research project “Finding out what works best in the measures taken against right-wing extremism in Switzerland” intends to improve our understanding of how measures against right-wing extremism can best be evaluated. The project aims on the one hand to develop guidelines for the evaluation of measures to combat right-wing extremism and on the other hand to provide information on which measures taken against right-wing extremism are effective.

In order to reach this goals, the research project consists of the following elements: The first step will be to collect and synthesize information from existing evaluations of programmes against right-wing extremism or similar measures in Switzerland and abroad. A meta-evaluation will be conducted as part of this synthesis to provide knowledge about how the effectiveness of measures against right-wing extremism can be assessed. The second step will be to create an inventory of successful projects by surveying organizations active in employing measures against right-wing extremism in Switzerland. Third, using the Delphi technique, experts will assess the success of the measures contained in the inventory. The objective here will be to select two best practice examples from each of three areas (police and penal law, youth work, school) as well as two examples from other fields. The success conditions for each of these eight best practices will be worked out in detailed evaluative case studies in a fourth step. These eight evaluative studies will be compared systematically in a fifth step, primarily to provide evidence about the generalizability of the criteria for success. On the basis of the results from the preceding steps, the sixth and final step will be to provide recommendations for the future design of (1) measures against right-wing extremism in Switzerland as well as (2) evaluation of the effectiveness of these measures.

The present paper is focusing on the meta-evaluation conducted as part of the evaluation synthesis in the first research step. First we will provide a brief overview of the procedures followed within the meta-evaluation, including a short description of the Evaluation Standards of SEVAL, the Swiss Evaluation Society, which we used as assessment criteria. Then, measures taken against right-wing extremism, their intervention fields and country of origin will be presented. Afterwards the search of evaluation studies and literature is shown. It will be demonstrated how these evaluations have been traced and what distinctive features they were made of. The meta-evaluation process and its findings are presented in a further step. In the final section, the consequences for the evaluation design of measures against right-wing extremism are discussed.

Approach and instruments

The term meta-evaluation is worth a definition, since there are several understandings of the term used in the literature. Meta-evaluation means here the evaluation of one or more evaluation studies – in contrast to an evaluation synthesis (summarizing the findings of several evaluation studies qualitatively), or a meta-analysis (consisting of a quantitative re-analysis of findings of several existing evaluation studies). Therefore the objects of our investigation are evaluation studies of measures against right-wing extremism. To determine the value of an evaluation, we have to rely on assessment criteria. How we selected the evaluation to be investigated will be described in the following section. We will turn here to the topic, how the assessment of each study was conducted.

As is true for evaluations themselves, meta-evaluations can be fashioned in quite different ways: as self-evaluations or heteronomous evaluations, executed internally or externally, or they can fulfill formative or summative functions. In other words, the meta-evaluative approach is multifunctional, as is the evaluation approach in general. It can serve specific functions, such as to provide quality assurance, control, or learning, but it also can be conceptualized as a part of the evaluation project or be an independent endeavor. It can be initiated by the evaluators themselves, the person or agency giving out the evaluation, or by any other stakeholder. Its implementation can include or exclude the parties involved in the evaluation process. The meta-evaluation can try to reach a common understanding of the assessment criteria used within the meta-evaluation or it can define the relevant criteria based on other values. Our meta-evaluative study in the field of measures against right-wing extremism is in this taxonomy an, ex-post, heteronomous, external meta-evaluation with a summative function.

The cornerstone of each evaluation is the choice of assessment criteria for valuing the worth or merit of the evaluation object (the evaluandum), in our case the evaluation studies. One can differentiate the demands made on evaluation according to two different assessment criteria, namely internal and external, and use them to assess the quality of an evaluation. Table 1 provides an overview of the respective reference levels with their corresponding assessment bases.

Table 1: Reference Levels for Evaluation

<i>Level</i>	<i>Object</i>	<i>Basis for Internal Assessment Criteria</i>	<i>Basis for External Assessment Criteria</i>
<i>First Order Constructs</i>	Programme, Project, Measure, etc.	Objectives of the Programme, Project, etc.	Social Significance of the Programme
<i>Second Order Constructs</i>	Evaluation	Objectives of the Evaluation	Evaluation Theory and Methodology
<i>Third Order Constructs</i>	Meta-evaluation	Objectives of the Meta-evaluation

An evaluator can thus assess a programme (or another evaluandum; a first order construct) based on the programme objectives (internal assessment criteria). But an evaluator can also assess a programme based on the social significance of the programme (external assessment criteria). The same reasoning can be applied to an evaluation as a second order construct, where the relevant assessment criteria would therefore first be the objectives of the evaluation (internal assessment criteria) and second general evaluation theory and methodology (external assessment criteria). By analogy, this argument can be pursued at the third (or subsequent) order.

The relevant point here is that if we want to investigate the quality of an evaluation, we essentially have two possible assessment criteria, the internal and the external. In our meta-evaluation in the field of measures against right-wing extremism we focused on an assessment by external criteria for evaluations, that is, on evaluation theory and methodology. For a prescription of quality in evaluation theory and methodology we selected so called evaluation standards. Evaluation standards allow for a more precise formulation of the external assessment criteria. As our research project is intending to provide results for the Swiss context, we used the Evaluation Standards of SEVAL, the Swiss Evaluation Society (SEVAL Standards)².

The SEVAL Standards, following the Program Evaluation Standards developed by the Joint Committee on Standards for Educational Evaluation (Joint Committee 1994), are based on the

² see www.seval.ch for a full presentation of the SEVAL Standards in English, French, and German. Since we have assessed not only evaluation studies from Switzerland, the selection of the SEVAL Standards as assessment criteria was not straightforward. Instead, it would have made sense, to use context sensitive assessment criteria. In order to maintain high homogeneity and to take into account the goal of the project focused on the Swiss context, we decided to choose the SEVAL Standards for all evaluation studies, wherever they come from. This does not at all imply that we assume a universal scope for the SEVAL Standards.

premise that an evaluation should at once be useful, feasible, proper, and accurate so as to fulfill the demands placed on it: good evaluations should therefore demonstrate all these characteristics. To make these category characteristics more tangible, the SEVAL Standards are subdivided into 27 individual Standards that fall into one of the four larger categories.

Figure 1: SEVAL Standards: Overview

U Utility

The utility standards guarantee that an evaluation is oriented to the information needs of the intended users of the evaluation.

U1 Identifying Stakeholders

Those persons participating in, and affected by, an evaluation are identified in order that their interests and needs can be taken into account.

U2 Clarifying the Objectives of the Evaluation

All persons who are involved in an evaluation will ensure that the objectives of the evaluation are clear to all stakeholders.

U3 Credibility

Those who conduct evaluations are both competent and trustworthy; this will help ensure the results an evaluation reaches are accorded the highest degree of acceptance and credibility.

U4 Scope and Selection of Information

The scope and selection of the information that has been collected makes it possible to ask pertinent questions about the object of the evaluation. Such scope and selection also takes into account the interests and needs of the parties commissioning the evaluation, as well as other stakeholders.

U5 Transparency of Value Judgments

The underlying reasoning and points of view upon which an interpretation of evaluation results rests are described in such a manner that the bases for the value judgments are clear.

U6 Comprehensiveness and Clarity in Reporting

Evaluation reports describe the object of evaluation, including its context, goals, questions posed, and procedures used, as well as the findings reached in the evaluation – in such a manner that the most pertinent information is available and readily comprehensible.

U7 Timely Reporting

Significant interim results, as well as final reports, are made available to the intended users such that they can be utilized in a timely manner.

U8 Evaluation Impact

The planning, execution, and presentation of an evaluation encourage stakeholders both to follow the evaluation process and to use the evaluation.

F Feasibility

The feasibility standards ensure that an evaluation is conducted in a realistic, well-considered, diplomatic and cost-conscious manner.

F1 Practical Procedures

Evaluation procedures are designed such that the information needed is collected without unduly disrupting the object of the evaluation or the evaluation itself.

F2 Anticipating Political Viability

The various positions of the different interests involved are taken into account in planning and carrying out an evaluation in order to win their cooperation and discourage possible efforts by one or another group to limit evaluation activities or distort or misuse the results.

F3 Cost Effectiveness

Evaluations produce information of a value that justifies the cost of producing them.

P Propriety

The propriety standards ensure that an evaluation is carried out in a legal and ethical manner and that the welfare of the stakeholders is given due attention.

P1 Formal Written Agreement

The duties of the parties who agree to conduct an evaluation (specifying what, how, by whom, and when what is to be done) are set forth in a written agreement in order to obligate the contracting parties to fulfil all the agreed upon conditions, or if not, to renegotiate the agreement.

P2 Ensuring Individual Rights and Well-Being

Evaluations are planned and executed in such a manner as to protect and respect the rights and well-being of individuals.

P3 Respecting Human Dignity

Evaluations are structured in such a manner that contacts between participants are marked by mutual respect.

P4 Complete and Balanced Assessment

Evaluations are complete and balanced when they assess and present the strengths and weaknesses that exist in the object being evaluated, in a manner that strengths can be built on and problem areas addressed.

P5 Making Findings Available

The parties who contract to an evaluation ensure that its results are made available to all potentially affected persons, as well as to all other who have a legitimate claim to receive them.

P6 Declaring Conflicts of Interest

Conflicts of interest are addressed openly and honestly so that they compromise the evaluation processes and results as little as possible.

A Accuracy

The accuracy standards ensure that an evaluation produces and disseminates valid and usable information.

A1 Precise Description of the Object of Evaluation

The object of an evaluation is to be clearly and precisely described, documented, and unambiguously identified.

A2 Analyzing the Context

The influences of the context on the object of evaluation are identified.

A3 Precise Description of Goals, Questions, and Procedures

The goals pursued, questions asked, and procedures used in the evaluation are sufficiently precisely described and documented that they can be identified as well as assessed.

A4 Trustworthy Sources of Information

The sources of information used in an evaluation are sufficiently precisely described that their adequacy can be assessed.

A5 Valid and Reliable Information

To ensure the validity and reliability of the interpretation, it is necessary to select, develop, and employ procedures for that given purpose.

A6 Systematic Checking for Errors

The information collected, analyzed, and presented in an evaluation is systematically checked for errors.

A7 Qualitative and Quantitative Analysis

Qualitative and quantitative information are systematically and appropriately analyzed in an evaluation, in a manner that the questions posed by the evaluation can actually be answered.

A8 Substantiated Conclusions

The conclusions reached by an evaluation are explicitly substantiated in such a manner that stakeholders can comprehend and judge them.

A9 Neutral Reporting

Reporting is free from distortion through personal feelings or preferences on the part of any party to the evaluation; evaluation reports present results in a neutral manner.

A10 Meta-evaluation

The evaluation itself will be evaluated on the basis of existing (or other relevant) Standards such that the evaluation is appropriately executed, and so that stakeholders can, in the end, assess the evaluation's strengths and weaknesses.

Since their establishment in 2000, the SEVAL Standards were used in different settings as assessment scheme for meta-evaluations (for example Kuffner 2000; Läubli Loud 2003; Lulofs/Arentsen 2001; Nideröst 2001; Simon 2000; Widmer et al. 2001; Wüest-Rudin 2002; compare Widmer 2004).

In the meta-evaluation in the field of measures against right-wing extremism, we assessed each evaluation study along each of the 27 standards. For reasons of economy, we were only able to consider the final reports of the evaluation studies as source of information. The inclusion of additional information would be desirable since many of the aspects of an evaluation (process as well as product) can not be sufficiently investigated relying on the final report only. In a first step we have coded the evaluation reports with the ATLAS.ti 4.2 software along the SEVAL Standards. Each text passage, sentence or word was assigned to the corresponding SEVAL Standard(s). After having completed the coding process, the information gathered for each criteria was edited. Then the information obtained was assessed in comparison with the requirements imposed by every single SEVAL Standard. This assessment concluded with classifying each evaluation study in respect to each standard as 'standard was

addressed’, ‘standard was partially addressed’, ‘standard was not addressed’, and ‘unable to judge’.

Measures against right-wing extremism and their evaluation

Searching measures and evaluations

In order to get informed about evaluation practices in the field of measures against right-wing extremism and to have a broad sample of evaluation studies as basis for the selection of studies to be assessed in the meta-evaluation, we conducted a broad survey of measures against right-wing extremism. Furthermore, due to existing knowledge (Pingel/Rieker 2003) – there exists only a few reports measuring achievement or failure of measures against right-wing extremism – we extended the search of literature and evaluations in the field of right-wing extremism for those also treating with: right-wing radicalism, xenophobia, racism, anti-Semitism, violence, mobbing, discrimination, prejudices, etc.

In addition to the search of literature and evaluations in journals, newspapers, books and the World Wide Web, several institutions and persons such as experts, administration departments, clubs, foundations, public social aid, schools, government departments in Germany, Switzerland, Great Britain, France, Austria, Netherlands, Belgium, Spain and Australia have been inquired about existing projects. More precisely we were asking for evaluations of projects concerning the issues right-wing extremism, right-wing radicalism, xenophobia, racism, anti-Semitism, violence, discrimination and prejudices. In the case of the US, we restricted the search to publications as journal articles and books. Table 2 gives an overview, how the letters of inquiry are distributed among the select countries.

Table 2: Survey of measures and evaluation and selection process

	<i>Projects in the survey</i>	<i>Responses to the survey (response rate in %)</i>	<i>Responses with evaluation (proportion of responses in %)</i>	<i>Additional studies included</i>	<i>Total number of evaluation studies</i>	<i>Studies in the meta-evaluation</i>
Germany	204	106 (52)	33 (31)	19	52	4
Switzerland	101	68 (67)	13 (19)	0	13	3
UK	95	17 (18)	2 (12)	3	5	1
France	58	12 (21)	0 (0)	0	0	0
Austria	42	18 (43)	0 (0)	0	0	0
Netherlands	30	5 (17)	0 (0)	3	3	0
Belgium	29	7 (24)	0 (0)	0	0	0
Spain	25	1 (4)	0 (0)	0	0	0
Australia	1	1 (100)	1 (100)	0	1	1
USA	-	-	-	23	23	2
others (as EU)	4	3 (75)	0 (0)	4	4	1
Total	589	238 (40)	49 (21)	52	101	12

For various reasons, these data have to be interpreted with caution:

- First, the focus of our survey is biased towards certain countries (as Germany and Switzerland). Therefore the number of projects is not representative and gives not a reliable picture of activities in the domain of measures against right-wing extremism.
- Second, the response rate could be influenced by influences out of control.
- Third, we can assume that the proportion of projects announcing an existing evaluation is overestimated in general, since these projects tend to answer more frequently to our survey.

Although some restrictions have to be taken into account, the results of the survey show as well some remarkable findings:

- In the first place the evaluation activities in the field of measures against right-wing extremism, in contrast to our expectations, seem to be quite frequent in Germany and Switzerland relative to the situation in Austria, Belgium, France, the Netherlands and Spain. The UK has a medium position in this respect. The results - especially those in the cases of Germany and Switzerland – are not conforming to the picture sketched by various country comparisons in evaluation in general (see for example Rist 1990, Furubo et al. 2002, or Leeuw 2004).
- Secondly, measures against right-wing extremism are widespread in Germany, although the picture is slightly flawed by our focus.

- Third, the total number of evaluation studies identified – although this number is boosted by the fact that we have included also measures against racism etc. (see below) – is reasonable high, although our survey has not included countries outside Europe (especially the US and Australia) in a systematic way.

The 101 evaluation studies are very different, they vary in various dimensions. Table 3 shows some of the diversity:

Table 3: Some descriptive characteristics of the evaluation studies:

Characteristics	Categories	GE	USA	CH	UK	NL	AUS	var.	Sum
evaluation report contains methodical report	yes	25	15	5	3	2	1	3	54
	no	27	8	8	2	1	-	1	47
type of measures evaluated	right-wing extremism	18	-	6	-	-	-	-	24
	racism, discrimination, etc.	33	21	7	5	3	1	4	74
	NA	1	2	-	-	-	-	-	3
intervention fields	school	21	14	-	2	-	-	2	39
	youth work	7	4	5	2	-	-	-	18
	police and judiciary	4	-	4	-	1	-	1	10
	various	14	4	4	1	2	1	1	27
	NA	6	1	-	-	-	-	-	7
type of measures	information/prevention	11	17	6	2	-	-	1	37
	further education	20	1	5	1	1	1	-	29
	various	21	1	2	2	2	-	3	31
	NA	-	4	-	-	-	-	-	4
timeframe	2004	1	-	2	-	-	-	-	3
	2003	15	-	5	-	-	-	1	21
	2002	10	-	2	-	-	-	-	12
	2001	11	-	2	1	-	-	1	15
	2000	3	1	1	-	-	1	1	7
	1999	1	11	1	-	-	-	1	14
	1998	3	2	-	1	1	-	-	7
	1997	4	1	-	-	-	-	-	5
	1996	-	1	-	1	1	-	-	3
	1990-1995	1	4	-	2	1	-	-	8
	1980-1990	-	3	-	-	-	-	-	3
	NA	3	-	-	-	-	-	-	3

*NA: not available or not appropriate

About half of the evaluation reports contain no report about the methods applied. Only one fourth of the evaluation studies under investigation are concerned with measures against right-wing extremism. The rest is dealing with measures in the fields of racism, discrimination, violence, integration and so on. Many evaluations come out of the field of education. Information or prevention and further education are the most common types of measures. One half of the evaluations were conducted in 2001 or later.

Germany

The most important reason why so many projects (52) have been accomplished in Germany can be found in its historical past. The intervention fields of more than half of these projects have been schools and youth work. About 20 percent can be assigned to further education and three projects to the penal system. The rest of the German projects can be attributed to other intervention fields, e.g. family, public etc. About 30 percent of the projects pertain to measures against right-wing extremism. The rest concerns measures against racism, anti-Semitism, xenophobia and others. The nature of measures has been evenly distributed in the fields of political education, different trainings (e.g. trainings to deal with aggressions, trainings to learn how to deal with conflicts), promotion of social and interethnic contacts and other fields like change in law or hotlines.

The evaluation reports also differ in their quality and length. Seven evaluations were shorter than 10 pages. Most of the reports were between 10 and 30 pages long. About 20 percent had an amount of 40 to 100 pages and only a few evaluations had more than 100 pages. For the most part, evaluations were formally correct. A majority of evaluations has been published between the year 2001 and 2004. Only little more than ten percent of the evaluations have been published before 2001. Some evaluations have been published as power point presentations, others as papers or books.

About half of the evaluations provide a methodical report. Most of the German projects have been evaluated externally. In many cases the course of project has been evaluated, but about twenty percent of the evaluation have taken into account the effects of it. A few reports evaluated both the process and the effects. In about ten evaluations observation was used as method for data collection. Others have used interviews or surveys by (standardised) questionnaires. In some cases, the type of data collection procedures applied was not evident in the reports.

Switzerland

The Swiss projects can be arranged in three different intervention fields. They are evenly distributed among the fields youth work, public and police as well as judiciary. In the intervention field of youth work, one particularly beard reference to theatres. Hotlines, magazines or local poster campaigns are measures, taken in public. The police and judiciary fight against right-wing extremism by auditing those being suspected, changing laws and preventing different groups from having confrontations. The measures observed can approximately be assigned to the same degree to measures against right-wing extremism and other measures, e.g. against racism, anti-Semitism and xenophobia.

With regard to evaluations, they were made by external evaluators. Furthermore these evaluations are either very short (less than ten pages) or large (more than 50 pages). Only a few studies are based upon surveys. More often, the analysis is based on observations and interviews with the project managers in charge. The larger part of the evaluations are of recent date, what means they have been published between 2002 and 2004. Little more than half the evaluations were evaluating the project process. However, two evaluations were looking for the long-term effects of these projects.

United States of America

Most of the American projects have been conducted in schools. Another intervention field chosen was the individual and intergroup level. A few studies have been done in the fields of family, public or community. More than half the projects tried to decrease prejudices through different trainings, intergroup contacts, discussions or curricula. Most of the remaining projects aimed at a change in attitude by supporting pro-social behaviour or tolerance.

All studies collected from US got published in journals. Therefore they mostly are between ten and thirty pages long. The available evaluation studies have been published between the year 1983 and 2000. Some of these figured out to be meta-evaluations. Others were summaries of different projects. A few projects have been evaluated externally by using surveys. For the most part the effects of the projects mattered. Several evaluation studies were measuring both short- and long-term effects.

United Kingdom, Netherlands, Canada, Australia and the European Union

The projects coming from the United Kingdom, Netherlands, Canada, Australia and the European Union are very different both, in the kind of intervention fields and in the measures taken against racism, anti-Semitism, violence or xenophobia. None of these projects showed measures taken against right-wing extremism.

These projects have been conducted in public service companies, football stadiums, schools or on an individual level. The measures have been as different as the intervention fields. Some measures tried to support intergroup contact, whereas others tried to support the integration of ethnic minorities or offered further education of enforced police controls.

The years of publication of these evaluations also differ compared to the other countries discussed above. They ranged between 1983 and 2003. To some extent the evaluation studies have been published in journals, others as papers, or as doctoral thesis. These evaluations are

based on primary or secondary analyses. Mostly they were looking at the effects of the projects both in the short and long run.

Selecting evaluations for inclusion in the meta-evaluation

To select twelve evaluation studies for inclusion in the meta-evaluation, we first defined certain mandatory requirements, namely enough information density in the final reports. Therefore, evaluation reports with less than three pages were excluded without further consideration. The further screening proceeded in three steps. In every particular step the project evaluations have been scrutinized related to specific dimensions.

First, all available evaluations have been registered by means of the developed pattern. This pattern contained both the selected criteria and authors as well as sponsorship. The first step simply consisted in finding out whether the dimension can be measured with the evaluation report only. After characterizing every evaluation in a standardized pattern we selected in a second step those evaluation reports, which were extensive enough to be meta-evaluated.

In consideration of the fact that we focus on right-wing extremism, most of the evaluations chosen for meta-evaluation should deal with this issue. A relative high proportion of evaluations have been received from Germany. Not only we wanted to make a selection of German evaluations but we also selected evaluations from other countries. A matter of particular interest was to select evaluations carried out in Switzerland because on further steps our research will turn its attention to this country. We considered evaluations of different intervention- and practice fields.

Applying the above criteria, about thirty evaluations were retained. From this thirty evaluation studies twelve have been selected to be meta-evaluated. In this last selection step we paid attention to select both process- and effect-oriented evaluations. Moreover, we were considering a balanced distribution of the evaluations selected in respect to origin, intervention type, and field of intervention. Furthermore, the selected evaluations should be as recent as possible, since we have the intention to reflect current practices.

Another two evaluations have been kept as “reserve” to replace studies in the sample for the case that significant problems occur during the meta-evaluation process.

The following twelve studies were finally selected for the meta-evaluation:

1) Students qualification and scholar-multipliers

Hupfeld, Jörg (2004): *Wissenschaftliche Begleitforschung: Schülerqualifikation als Präventionsstrategie und Schülermultiplikatoren gegen Rechtsextremismus*. arpos Institut e.V. – not published

This project made a contribution to the prevention of right-wing extremism, xenophobia and anti-Semitism in Germany. The project offered different measures (e.g. construction of an internet page, exchange of information) to adolescents.

The project has been evaluated externally. The effect of the project has been evaluated due to a quantitative questionnaire. The process of the project has been explored by qualitative interviews and observations.

2) Campaign ‘Facts against Prejudices’

Artho, Jürg (2003): *Evaluation der Kampagne ‚Tatsachen gegen Vorurteile‘*. Sozialforschungsstelle Zürich. – not published

The Swiss cantons Basel-Stadt und Basel-Landschaft launched a poster campaign, advertisements in newspapers and a migration magazine to prevent racism, anti-Semitism and xenophobia. To measure the effect of the campaign an external evaluation has been done.

To answer the evaluation questions (e.g. How strong was the percipience of the campaign? Who gathered the contains of the campaign? How strong was the effect of the campaign?) evaluators applied a written survey.

3) Exploring Political Tolerance with Adolescents

Avery, Patricia G./Brid, Karen/Johnstone, Sandra/Sullivan, John L./Thalhammer, Kristina (1992): *Exploring Political Tolerance with Adolescents. Theory and Research in Social Education*. Fall, 1992, Volume XX, Number 4, pp. 386-420.

The Curriculum “Tolerance for Diversity of Beliefs” has been applied in US-classes to increase political tolerance.

The aim of the evaluation was to detect whether the political tolerance increases if adolescents were taught with the curriculum. The authors acquired the findings with questionnaires distributed to each student.

4) FURD (Football Unites, Racism Divides)

Bradbury, Steven (2001): *Football Unites, Racism Divides. An evaluation report 1998-2000*. Sir Norman Chester Centre for Football Research – University of Leicester. – not published

The English project FURD aims “(...) to ensure that everyone who plays or watches football can do so without fear of racial abuse and harassment, in either a verbal or a physical form, and to increase the participation of people from ethnic minorities in football”

(www.furd.org). The evaluation had to provide a critical assessment of the process made by the FURD project. The evaluator applied different sources e.g. interviews, analyses reports, observations or other documents.

5) Evaluation of the European Monitoring Centre on Racism and Xenophobia (EUMC)

Centre for Strategy & Evaluation Services (2002): *Abschlussbericht über die Evaluierung der Europäischen Stelle zur Beobachtung von Rassismus und Fremdenfeindlichkeit*.
Online: http://europa.eu.int/comm/employment_social/fundamental_rights/pdf/origin/eumc_eval2002_de.pdf (2004-09-10)

This evaluation was undertaken for the European Commission between September 2001 and March 2002. It takes into account the developments of the EUMC from the setting up until the end of 2001. The evaluators primarily based the survey on EUMC publications, but also a written inquiry was posted.

6) Prevention of and combat against xenophobic, racist and mightily attitudes

Eser, Miryam (1999): *Prävention und Bekämpfung fremdenfeindlicher, rassistischer und gewaltaffiner Einstellungen – Evaluationsstudie eines einstellungsverändernden Projekts mit BerufsschülerInnen. Abhandlung zur Erlangung der Doktorwürde der Philosophischen Fakultät I der Universität Zürich*. Bern: Edition Soziothek.

This project was being supported by the Swiss National Science Foundation. The project aimed to prevent and reduce intolerant attitudes and prejudices. A curriculum has been applied to assist tolerance. The evaluation is a dissertation. Process and effect of the project have been evaluated due three written inquiries.

7) Adolescents – racism – attitudes

Fabian, Carlo/Gutmann, Renate (2004): *Jugendliche – Rassismus – Einstellungen. Die Bedeutung von personalen und sozialen Ressourcen – Analysen am Beispiel des Forumtheaters CRASH!* – not published

For about sixteen years a Swiss troupe arranges different projects to call attention to racism and disrespect. The evaluation study evaluated the play CRASH!. The play has been developed for children and adolescents. The evaluation wanted to figure out the preventive effect of the project.

8) Violence Prevention Curriculum Among Children in Elementary School

Grossman, D.C. et al. (1997): *A Randomized Controlled Trail of a Violence Prevention Curriculum among Elementary School Children*. In: Journal of the American Medical Association. Jg. 277: 1605 – 1611.

The violence preventing curriculum “Second Step: A Violence Preventing Curriculum, Grades 1-3” has been applied in different classes in the school district of King County, Washington. The curriculum tried to increase neutral and pro-social behaviour.

The evaluators tried to find out, how effective the curriculum “Second Step: A Violence Preventing Curriculum, Grades 1-3” was. They used observation reports done by teachers, parents and observers.

9) Stereotypes, Prejudices, and Prejudice Reduction

Hill, Miriam E. (2000): *Stereotype Change and Prejudice Reduction: Short- and Long-term Evaluation of a Cross-cultural Awareness Programme*. – not published

The Cross-Cultural Awareness Programme has been developed by the Courts Administration Authority of South Australia (CAA) to reduce prejudice compared to Aboriginal Australians. In a course CAA employees were informed about the history of Aboriginal Australians. Furthermore they had to deal with their own prejudices.

The evaluation is a doctoral thesis. With written questionnaires the author tried to find out, if the programme has the effect of increasing endorsement of sociological stereotypes between different points in time.

10) PAT (Pro Acceptance and Tolerance – a multiplier project for instructors)

Institut für Arbeitsmarktforschung und Jugendberufsbildung (2004): *Förderung der Handlungskompetenz im Umgang mit Gewalt und Fremdenfeindlichkeit in der Berufsausbildung – eine realistische Zielsetzung von XENOS-Projekten?* – not published

This German projects comprises measures such as seminars and workshops in order to get knowledge on decision-making and responsibility across adolescents dealing with aggressiveness, violence and right-wing extremism during their formation.

The evaluation was done externally and in a formative way. Qualitative as well as quantitative data were collected on participants, their institutions and the course as results of seminars.

11) Cinema for tolerance

Institut für Kino und Filmkultur (2003): *Ergebnisbericht und Evaluationsbericht zum Projekt Kino für Toleranz 2003 (als Fortführung des Projekts Kino gegen Gewalt)*. – not published

The action implemented in the context of this German project is nationwide cinema shows which are accompanied by pedagogues. The main audience being reached are students and their teachers. This project has been evaluated since the beginning of October until the end of November 2003. The survey instrument chosen was a written interview. On the one hand, the study evaluated the effectiveness and thematic relevance for students purposes and otherwise the impact of influence,

pedagogical preparation and post processing and assessment of the project idea as show for teachers purposes.

12) CAT (Creative and Active Training)

GESOMED – Gesellschaft für sozialwissenschaftliche Forschung in der Medizin mbH (2003): *CAT – Creative and Active Training*. – not published
Online: <http://www.powerforpeace.de/documents/cat-eval-2003.pdf> (2004-09-10).

This German project can be classified in the range of local networks. CAT is a training for students and juveniles to establish understanding, trust and to build up a sense of community. More specific, it is an anti-violence training that acts on the (potential) offender. The underlying concept is based on the Alternatives to Violence Project (AVP), U.S.A. The seminars have been held in the following institutions: prisons, supporting school, secondary school and vocational school The process and effect of these trainings were taken into account. The evaluators developed therefore multiple course-specific questionnaires.

Results of the meta-evaluation

Along the four standard groups of the SEVAL Standards the following results have been accomplished.

Utility

Overall, the standards criteria for this subject group were fulfilled to a satisfactorily extent. The identification of stakeholders generally was in accordance with the demands of the standard. Almost without exception those persons, groups and/or institutions participating in or affected by an evaluation have been specified. However in many cases not much value was set on covering the interests and needs of the stakeholders. With no exception, every evaluation clarified its objectives in a comprehensive way. In the majority of the cases information lacking about evaluators did not permit an appropriate assessment of the ‘Credibility’ Standard. Even additional information collected from the World Wide Web did not help much to enlighten the competency and trustworthiness of evaluators. ‘Scope and Selection of Information’ either fully or partially met the criteria of this Standard. In many cases the appendix was incomplete or too short. Drawing a positive balance of the ‘Transparency of Value Judgments’ Standard, the evaluations either fully or partially met the criteria. In need of improve-

ment are the descriptions of underlying approaches in order to assign value to the gained information. In general, the relevant information (e.g. programme, context, aims, procedures and results of the evaluations) were available and could be understood easily. No evaluation informed about time frames. Sometimes the period of time between the end of the project and the completion of the evaluation study was so short, that the timeliness could be assumed in favour of the evaluators. With one exception, it was not apparent if any evaluation showed an impact in practice.

Feasibility

The procedures chosen were not always appropriate. Sometimes the data required could have been collected in a more practical way without any extra cost or effort. It seems like the consequences from employing specific methods or procedures have not been fully considered in advance. Therefore the results were often incomplete or little extensive. The ‘Political Viability’ and the ‘Cost Effectiveness’ could never be judged because the corresponding information was missing.

Propriety

In the majority of cases no statements could be made about the criteria ‘Formal Written Agreement’. Where it came to an assessment it was due to speculations. There is evidence to suggest that all the evaluations have been planned and executed in such a manner as to ensure individual rights and well-being of individuals. If they were secured during the planning and carrying out of an evaluation could not explicitly seen from the texts, but assumed. Due to missing information the Standards dealing with ‘Respecting Human Dignity’ and ‘Declaring Conflicts of Interest’ could not be judged. Strengths and weaknesses of the projects have been presented in the evaluation reports. The judgments made were complete and balanced. The findings of each evaluation were accessible for the public over different kinds of media. If not, they have been sent to us easily on request.

Accuracy

The object of evaluation has always been described, documented and identified in a at least adequate way. Sometimes in short evaluations this description has not been very detailed, but the evaluators made cross references to further information sources. The context analyses were of different quality. All evaluation studies share a well-done characterisations of samples and key stakeholders. Differences arose in including e.g. the political or social context. In general the goals and questions pursued were communicated in a clear and comprehensive

language. As mentioned before the documenting of procedures partially showed some weaknesses. As goals, questions and procedures were not described sufficiently precise, the process of evaluation was not always transparent and comprehensible. Beside evaluations which accurately described the sources of information used, others described these sources very inadequately or not at all. As a conclusion the latter could not be judged in order to their trustworthiness. The validity and reliability of the measurement and measuring instruments overall were guaranteed in a sufficient degree. Spelling mistakes indeed appeared but seldom. In some evaluations data have not been checked systematically (e.g. miscalculations, biased graphics etc.). We did not find indications of mistaken interpretations of the data collected. It was not easy judging the Standard ‘Qualitative and Quantitative Analysis’ as the methodology choices often were not clearly described. Furthermore no evaluation was accompanied by the complete data set. Also, the analyses of information repeatedly did not correspond to the measuring scales. The conclusions reached in the evaluations were explicitly justified and clearly described. Nearly always, we were not able to detect reporting distorted by personal feelings or preferences on the part of any party of evaluation. The conclusions were drawn in a neutral manner. By means of the documentation we could not find any meta-evaluations made in the framework of the studies under investigation.

To avoid misleading interpretation, we do not present the results of the meta-evaluation study by study. Otherwise, it would be possible to sum up the judgements received by each study and to rank them accordingly. This would not be appropriate, since the importance of each standard can vary from study to study. The following table is showing an overview for the results produced in the meta-evaluation. These results are not yet double-checked. Therefore they have a preliminary status.

Table 4: Results of the meta-evaluation (preliminary results)

Standard was ...		addressed	partially addressed	not addressed	unable to judge
Utility:					
U1	Identifying Stakeholders	6	6	-	-
U2	Clarifying the Objectives of the Evaluation	12	-	-	-
U3	Credibility	3	-	-	9
U4	Scope and Selection of Information	6	5	-	1
U5	Transparency of Value Judgments	6	5	1	-
U6	Comprehensiveness and Clarity in Reporting	6	6	-	-
U7	Timely Reporting	3	-	-	9
U8	Evaluation Impact	-	1	-	11
Feasibility:					
F1	Practical Procedures	6	3	1	2
F2	Anticipating Political Viability	-	-	-	12
F3	Cost Effectiveness	-	1	-	11
Propriety:					
P1	Formal Written Agreement	1	1	-	10
P2	Ensuring Individual Rights and Well-Being	12	-	-	-
P3	Respecting Human Dignity	3	-	-	9
P4	Complete and Balanced Assessment	11	-	-	1
P5	Making Findings Available	10	2	-	-
P6	Declaring Conflicts of Interest	2	-	-	10
Accuracy:					
A1	Precise Description of the Object of Evaluation	9	3	-	-
A2	Analyzing the Context	7	4	1	-
A3	Precise Description of Goals, Questions, Procedures	7	4	1	-
A4	Trustworthy Sources of Information	7	3	1	1
A5	Valid and Reliable Information	6	3	-	3
A6	Systematic Checking for Errors	8	3	-	1
A7	Qualitative and Quantitative Analysis	1	5	1	5
A8	Substantiated Conclusions	9	3	-	-
A9	Neutral Reporting	11	1	-	-
A10	Meta-Evaluation	-	-	-	12

The results vary from standard to standard. In terms of assessments, we can divide the standards roughly in three groups:

- U3, U7, U8, F2, F3, P1, P3, P6, A10: For these standards a judgment was in many cases not possible, since the information gathered in the meta-evaluation was insufficient to judge. Under the given economic restrictions and in order to include as many studies in our meta-evaluation as possible, we decided to consult the final reports only. Additional documents or interviews with evaluators or commissioners would improve the information base and would allow for an assessment more often.
- U2, P2, P4, P5, and A9: These standards were addressed in all or nearly all studies. These standards are respected mostly in an appropriate way and therefore, the practice observed fulfills the requirement imposed by the SEVAL Standards. In the case a judgement was possible, the standards in the Propriety group were addressed sufficiently in most cases.

- U1, U4, U5, U6, F1, A1, A2, A3, A4, A5, A6, A7, A8: These standards are only fully addressed in a part of the studies. Often, these standards are only partially, in rare cases not at all respected. Here we can locate possibilities for improvement of evaluation practice. No standards from the Feasibility group are included here (for two out of three Feasibility standards, an assessment was not possible), whereas many of the standards concerned belong to the Accuracy group. In addition, one half of the Utility standards were in one half of the studies only partially addressed.

Summary

As our survey has shown, there is a growing evaluation activity in the field of measures against right-wing extremism and similar fields as racism or violence. But only one quarter of the evaluation studies identified are directly dealing with measures against right-wing extremism. Most of the others have measures against racism, discrimination, and violence in their focus. Most of the measures evaluated are located in education. Rarer are studies in youth work or police and judiciary. More than one half of the evaluation reports at our disposal do not include a methodical reporting allowing to understand, how the evaluation was conducted. Overall, one fifth of the measures identified were evaluated. Surprisingly, Germany and Switzerland are the countries under investigation with the highest proportion of measures with an evaluation. Although 18 respectively twelve measures could be identified for Austria or France, not one single Austrian or French project was able to present an evaluation study.

Among the measures under investigation, the variety in terms of intervention field or type of intervention is large. The same holds true for the evaluations of these interventions. Some of them are small-scale practice oriented self-evaluations with a short, mainly results oriented report without any methodological transparency. On the other hand, there exist some large, very well documented studies, where an academic orientation is dominant. Most of the studies are in the middle of these two extremes.

For our meta-evaluation, we have selected twelve out of the 101 evaluation studies identified. For the selection process, we have taken into account various dimensions. For example, we have excluded all studies where the reporting was not providing enough information about the methods applied. In addition, we have chosen a selection of studies representing the big diversity in terms of evaluandum and evaluation approach.

The meta-evaluation has shown, that the quality of the studies under investigation is quite heterogenous. Whereas some of the studies address most of the standards to a high degree,

some of them are neglecting one or several of the standards partially or (in rare cases) fully. In general, the quality is quite good. However, there are opportunities for improvements as well. This holds true especially for some of the Utility and some of the Accuracy standards.

In our meta-evaluation we have applied the SEVAL Standards as assessment criteria. We did not take into account the individual objectives set within a single evaluation. Therefore our assessment covers only one part of the overall quality since it does not consider the internal assessment criteria. The SEVAL Standards have – again – proved to be an applicable and useful tool for the assessment of evaluation quality in a meta-evaluation. The biggest problem during the meta-evaluation was the lack of information available due to our decision to restrict the focus of the meta-evaluation to the evaluation reports only without considering other sources of information. As a consequence, a quality assessment was not possible for every standard.

Based on the empirical findings, we will proceed in our research project by formulation a first draft of guidelines for the evaluation of right-wing extremism in order to help people involved in the evaluation of such measures to improve evaluation quality.

References

- Furubo, Jan-Eric/Rist, Ray C./Sandahl, Rolf (eds.) (2002): *International Atlas of Evaluation*. New Brunswick: Transaction.
- Joint Committee on Standards for Educational Evaluation (ed.) (1994): *The Program Evaluation Standards*. Newbury Park: Sage.
- Kuffner, Astrid (2000): *Evaluation von Nachhaltigkeitsaspekten – Nachhaltige Evaluation?* Diploma Thesis: University of Vienna.
- Läubli Loud, Marlène M. (2003): *Setting Standards and Providing Guidelines: the means towards what end?* Paper submitted to Fifth European Conference on Evaluation of the Structural Funds “Challenges for Evaluation in and Enlarged Europe”. Budapest, 26/27 June 2003.
- Leeuw, Frans L. (2004): Evaluation in Europe. In: Stockmann, Reinhard (Hrsg.): *Evaluationsforschung*. 2. Auflage. Opladen: Leske + Budrich: 61-81.
- Lehmann, Luzia/Balthasar, Andreas (2004): *Quality Assessment of External Evaluation Reports Commissioned by the Swiss Agency for Development and Cooperation. A Case of Evaluation Standards Put to Practice*. Paper to be presented at the 6th conference of the European Evaluation Society (EES) in Berlin, September 30 – October 2, 2004.
- Lulofs, K./Arentsen M. (2001): *Improving Quality and Learning Performance of ‘Energie 2000’*. Bern: Bundesamt für Energie.
- Nideröst, Bruno (2002): *Erfolgsbedingungen für Evaluationen*. LeGes – Gesetzgebung & Evaluation 13(1): 39-55.
- Pingel, Andrea/Rieker, Peter (2003): Evaluation pädagogischer Praxis gegen Rechtsextremismus: Potenziale und Chancen. In: Von Berg, Heinz Lynen /Roth, Roland (Hrg.). *Massnahmen und Programme gegen Rechtsextremismus wissenschaftlich begleitet. Aufgaben, Konzepte und Erfahrungen*. Opladen: 103-18.
- Rist, Ray C. (ed.) (1990): *Program Evaluation and the Management of Government*. New Brunswick: Transaction.
- Simon, Christian (2000): *Meta-Evaluation einer ausseruniversitären Institution der Lehre*. Berne: Center for Science and Technology Studies.
- Widmer, Thomas (2000): *Evaluating Evaluations: Does the Swiss Practice Live up to The Program Evaluation Standards?* Evaluation Center Occasional Papers Series, Vol. 17. Kalamazoo, Western Michigan University: 67-80.
- Widmer, Thomas (2004): Instruments and Procedures for Assuring Evaluation Quality. A Swiss Perspective. In: Schwartz, Reuven /Mayne, John /Toulemonde, Jacques (eds.): *Assuring the Quality of Evaluative Information: Prospects and Pitfalls*. New Brunswick: Transaction (forthcoming).
- Widmer, Thomas/Rüegg, Erwin/Neuenschwander, Peter (2001): *Stand und Aussichten der Evaluation beim Bund (EvalBund)*. Bern: Schweizerische Bundeskanzlei.
- Wüest-Rudin, Daniel (2002): *Evaluation von Reformen der öffentlichen Verwaltung: Bedeutung und Erfolgsfaktoren*. LeGes – Gesetzgebung & Evaluation 13(1): 57-82.