



**University of
Zurich**^{UZH}

**Zurich Open Repository and
Archive**

University of Zurich
University Library
Strickhofstrasse 39
CH-8057 Zurich
www.zora.uzh.ch

Year: 2007

Evaluating Measures Taken Against Right-Wing Extremism

Widmer, Thomas ; Blaser, Cornelia ; Falk, Chantal

DOI: <https://doi.org/10.1177/1356389007075225>

Posted at the Zurich Open Repository and Archive, University of Zurich

ZORA URL: <https://doi.org/10.5167/uzh-66039>

Journal Article

Accepted Version

Originally published at:

Widmer, Thomas; Blaser, Cornelia; Falk, Chantal (2007). Evaluating Measures Taken Against Right-Wing Extremism. *Evaluation*, 13(2):221-239.

DOI: <https://doi.org/10.1177/1356389007075225>

Evaluating Measures Taken Against Right-Wing Extremism

Thomas Widmer, Cornelia Blaser and Chantal Falk

Accepted for publication in 'Evaluation', Volume 13, Issue 2 (April 2007)

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 presents the results of the meta-evaluation in the field of measures against right-wing extremism in areas including education, social work and others.

The meta-evaluation follows the professional evaluation standards set by the Swiss Evaluation Society (SEVAL Standards). The analyzing 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 of a systematic analysis of twelve evaluation reports. 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 presents evidence about the usefulness of the SEVAL Standards in an applied setting. In this spirit this paper provides background information on both, evaluation of measures against right-wing extremism and the applicability of evaluation standards.

Keywords:

measures against right-wing extremism, evaluation standards, meta-evaluation, Switzerland, SEVAL Standards

Running head:

Widmer et al: Evaluating Measures Taken Against Right-Wing Extremism

Author biographies:

THOMAS WIDMER is a lecturer and the head of the Research Area ‘Policy Analysis & Evaluation’, Department of Political Science, University of Zurich.

Please send correspondence to: Department of Political Science, University of Zurich, Seilergraben 53, CH-8001 Zurich, Switzerland. [e-mail: thow@pwi.unizh.ch]

CORNELIA BLASER is a research fellow in the Research Area ‘Policy Analysis & Evaluation’, Department of Political Science, University of Zurich.

Please send correspondence to: Department of Political Science, University of Zurich, Seilergraben 53, CH-8001 Zurich, Switzerland. [e-mail: blaser@pwi.unizh.ch]

CHANTAL FALK is a research fellow in the Research Area ‘Policy Analysis & Evaluation’, Department of Political Science, University of Zurich

Please send correspondence to: Department of Political Science, University of Zurich, Seilergraben 53, CH-8001 Zurich, Switzerland. [e-mail: chanfa@access.unizh.ch]

Évaluation des mesures conçues pour lutter contre l'extrémisme de droite

Thomas Widmer, Cornelia Blaser et Chantal Falk

Résumé

Dans le cadre d'un projet de recherche plus large sur l'extrémisme de droite, ses causes et les mesures possibles à prendre à son encontre, nous examinons la qualité d'évaluations menées en Suisse, en Allemagne, aux Etats-Unis et dans d'autres pays dans le but de synthétiser leurs résultats. Cet article présente les résultats de la méta-évaluation des mesures contre l'extrémisme de droite dans les domaines tels que l'éducation, le travail social, etc.

La méta-évaluation est conduite selon les standards professionnels définis par la Société Suisse d'Évaluation (Standards SEVAL). L'analyse des évaluations sélectionnées dans le domaine des mesures prises contre l'extrémisme de droite nous permet d'apprécier la qualité de leurs résultats. La méta-évaluation consiste en une analyse systématique de douze rapports d'évaluation. Ceci nous permet d'élaborer une vision d'ensemble des pratiques courantes d'évaluation concernant les mesures prises contre l'extrémisme de droite dans quelques pays choisis. En outre, cet article démontre l'utilité des standards SEVAL dans un contexte spécifique. À cet égard, cet article livre des informations de fond sur deux éléments qui sont, d'une part, l'évaluation des mesures contre l'extrémisme de droite et d'autre part, l'applicabilité des standards d'évaluation.

Mots-clés:

Mesures contre l'extrémisme de droite, standards d'évaluation, méta-évaluation, Suisse, standards SEVAL

Introduction¹

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 Councilor Mrs. Ruth Metzler, formed a study group on 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 Program “Violence in Daily Life and Organized Crime” (NRP 40) to include the topic “Right-wing Extremism – Causes and Countermeasures” by adding the NRP 40plus². 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. The program centers around four research foci, 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 effectiveness 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.

Our research project, entitled “Finding out what works best in the measures taken against right-wing extremism in Switzerland”, intends to serve two purposes. First, it aims at improving our understanding of how measures against right-wing extremism can best be evaluated. As an outcome we will develop practical guidelines for the evaluation of such measures for practitioners of evaluation as well as for scholars who are concerned with right-wing extremism. The nature of right-wing extremism as an objective of investigation places specific demands on an evaluation that have constantly to be beard in mind. 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. In order to accomplish the first goal, the first research step

¹ An earlier version of this article has been presented at the EES conference in Berlin, September 30 – October 2, 2004. Financial support by the Swiss National Science Foundation (grant no. 404040-101065) is gratefully acknowledged. We would like to thank John Bendix for his collaboration as well as the two anonymous reviewers for their helpful comments.

² See www.nfp40plus.ch (in French and German only) for further information about this research programme.

consists of collecting and synthesizing information from existing evaluations of measures against right-wing extremism in Switzerland and abroad. This enables us to sketch a general picture of the current practice of evaluation of measures against right-wing extremism in selected countries. Second, a meta-evaluation of twelve evaluations of measures against right-wing extremism at focus will be conducted as part of this synthesis in order to assess the quality of these on the one hand and to provide knowledge about how the effectiveness of such measures can be evaluated on the other hand. The meta-evaluation follows the Evaluation Standards of SEVAL, the Swiss Evaluation Society. The discussion of the usefulness of the SEVAL Standards will contribute to evaluation theory and practice.

The second objective of our research project, which is not subject of this paper, is to provide information on which measures taken against right-wing extremism are effective. An evaluation synthesis, which is based on the selected evaluations, will provide evidence about the characteristics of the measures to combat right-wing extremism. In a next step, we will create an inventory of successful projects by surveying organizations active in employing measures against right-wing extremism in Switzerland. Second, 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 third step. These eight evaluative studies will be compared systematically in a fourth step, primarily to provide evidence about the generalizability of the criteria for success. On the basis of the results from the preceding steps, the fifth 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.

In a nutshell, the proceeding is comprised of six modules referring to subsequent stages in the research process.

Table 1: The six elements of the research project

<p><i>Module 1: Evaluation synthesis</i></p> <ul style="list-style-type: none"> • collecting studies • assessing evaluation quality • synthesizing findings 	<p><i>Module 2: Inventory of measures</i></p> <ul style="list-style-type: none"> • survey 	<p><i>Module 3: Panel of experts</i></p> <ul style="list-style-type: none"> • expert survey
<p><i>Module 4: Evaluative case studies</i></p> <ul style="list-style-type: none"> • conceptual work • case studies 1-8 	<p><i>Module 5: Comparison</i></p> <ul style="list-style-type: none"> • comparing 	<p><i>Module 6: Synthesis</i></p> <ul style="list-style-type: none"> • synthesizing

The present paper focuses on the first module of our research project and in particular on the meta-evaluation conducted as part of the evaluation synthesis. First we begin with a brief introduction in meta-evaluation and an overview of the procedures followed within the meta-evaluation, including a short description of the SEVAL Standards, which we used as assessment criteria. Then, measures taken against right-wing extremism, their intervention fields and country of origin will be presented. We then turn to an examination of the literature, showing how the various evaluation studies were traced and what their distinctive features are. 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. In addition we discuss in what respect evaluation standards were useful to assess the quality of evaluation studies.

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. 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 purposes. 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 in this taxonomy is 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 2 provides an overview of the respective reference levels with their corresponding assessment bases.

Table 2: 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>	Program, Project, Measure, etc.	Objectives of the Program, Project, etc.	Social Significance of the Program
<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 program (or another evaluandum; a first order construct) based on the program objectives (internal assessment criteria). But an evaluator can also assess a program based on the social significance of the program (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 ones. 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)³.

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 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

Utility	U1	Identifying Stakeholders
	U2	Clarifying the Objectives of the Evaluation
	U3	Credibility
	U4	Scope and Selection of Information
	U5	Transparency of Value Judgments
	U6	Comprehensiveness and Clarity in Reporting
	U7	Timely Reporting
	U8	Evaluation Impact
Feasibility	F1	Practical Procedures
	F2	Anticipating Political Viability
	F3	Cost Effectiveness
Propriety	P1	Formal Written Agreement
	P2	Ensuring Individual Rights and Well-Being
	P3	Respecting Human Dignity
	P4	Complete and Balanced Assessment
	P5	Making Findings Available
	P6	Declaring Conflicts of Interest
Accuracy	A1	Precise Description of the Object of Evaluation
	A2	Analyzing the Context
	A3	Precise Description of Goals, Questions, and Procedures
	A4	Trustworthy Sources of Information
	A5	Valid and Reliable Information
	A6	Systematic Checking for Errors
	A7	Qualitative and Quantitative Analysis
	A8	Substantiated Conclusions
	A9	Neutral Reporting
	A10	Meta-evaluation

³ 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. For a comparison of various standards-like documents, see Widmer 2005b.

Since their codification in 2000, the SEVAL Standards have been used in different settings as a means for assessing evaluation quality (for example Kuffner 2000; Läubli Loud 2004; Lehmann/Balthasar 2004; Lulofs/Arentsen 2001; Nideröst 2001; Wüest-Rudin 2002; compare Widmer 2005a).

In the meta-evaluation in the field of measures against right-wing extremism, we assessed each evaluation study along each of the 27 standards in qualitative case studies. The demands placed on an evaluation are based on maximum stakes. Therefore, it was neither a matter of ticking boxes or counting if some minimal standards have been fulfilled nor discharging checklists (Scriven 1991: 204-211, 230). Instead, our proceeding was oriented towards a deductive content analysis, which has been made transparent and producible for third parties by explaining the underlying rationale for each assessment. For reasons of research economy, we had to restrict our investigation to the final reports of the evaluation. 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 standard(s). After having completed the coding process, the information gathered for each criterion 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 with respect to each standard as ‘standard was addressed’, ‘standard was partially addressed’, ‘standard was not addressed’, and ‘unable to judge’. Each assessment has been double-checked in order to improve reliability.

Measures against right-wing extremism and their evaluation

Searching measures and evaluations

We conducted a broad survey of measures taken against right-wing extremism. Few reports (see Pingel/Rieker 2003) provide information about whether measures specifically taken against right-wing extremism were successes or failures, let alone measure them, so we extended our search to include studies conducted in related topics: right-wing radicalism, xenophobia, racism, violence, discrimination, prejudices, etc. This proceeding is in accordance with the definition of right-wing extremism we base our study on.⁴

⁴ Many definitions of right-wing extremism can be identified in the literature. We base our study on the widely accepted definition of Jaschke (1994) which perceives right-wing extremism as the attitudes, beliefs, views or behaviour of individuals or groups, which deny the basic claims of equality represented by equal rights due to racist or ethical reasons. Xenophobia for instance is conceptualised as a dimensional aspect of right-wing extremism. This view is supported by previous research, which sees right-wing extremism mainly affected by the

In addition to the search of literature and evaluations in journals, newspapers, books and the World Wide Web, along with consulting experts, we also enquired about existing evaluation studies among administrative departments, clubs, foundations, public social aid, schools and government departments in Germany, Switzerland, Great Britain, France, Austria, Netherlands, Belgium, Spain and Australia. In the case of the US, we restricted the search to publications as journal articles and books. Table 3 provides information about the scope of the survey and the response rate.

Table 3: Measures and evaluations in the survey and in the sample

	<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, some countries (as Germany and Switzerland) are overrepresented in our survey.
- Second, the response rate could be influenced by factors beyond our knowledge.
- 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:

social mid with its racist and xenophobic attitudes (Ahlheim 2003, Falter 1994, Steffens 2003). Henceforth we use the term measures against right-wing extremism in a broader sense.

- 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 – do not conform to the general image provided by inter country comparisons of evaluation activities (Rist 1990, Furubo et al. 2002, 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 vary along several dimensions as can be seen in Table 4.

Table 4: Some descriptive characteristics of the evaluation studies

Characteristics	Categories	GER	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 deals 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 educational measures (e.g. to deal with aggression, to learn how to deal with conflict), 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 were from 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 were published between 2001 and 2004. Only little more than ten percent of the evaluations were published before 2001. Some evaluations were published as transparencies, others as papers, brochures 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 20 percent of the evaluations have taken the effects of the measures into account. A few reports evaluated both the process and the effects. In 10 evaluations observation was used as method for data collection. Others have used interviews or surveys by (standardized) 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, there is particularly broad reference to the use of acting and role-playing. In measures directed towards the general public, hotlines, magazines or local poster campaigns are the instruments most often used. The police and judiciary fight against right-wing extremism by keeping those suspected under surveillance, changing laws and preventing different groups from having confrontations. Approximately half of the measures observed are measures against right-wing extremism in a narrow sense; the rest consists of related measures (e.g. racism, violence, xenophobia).

With regard to evaluations, they were made by external evaluators. Furthermore these evaluations are either very short (less than ten pages) or very long (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 is 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 inter-group 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 various educational activities, inter-group contacts, discussions or curricula. Most of the remaining projects aimed at a change in attitude by supporting pro-social behavior or tolerance.

All studies collected from US were published in journals. Therefore they mostly are between 10 and 30 pages long. The available evaluation studies were published between 1983 and 2000. Some of them proved to be meta-evaluations, while others summarized different studies. 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 in sensu stricto.

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 inter-group contact, whereas others tried to support the integration of ethnic minorities or offered further education or 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 theses. 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 the selected criteria and authors as well as sponsorship. The first step simply consisted in finding out whether the dimension can be assessed with the information available. 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 were found in Germany. Not only did we want 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 the subsequent steps in

our research focus more specifically to this country. We considered evaluations of different intervention fields.

Applying the above criteria, about thirty evaluations were retained. From these thirty evaluation studies, twelve were selected for meta-evaluation. In this last selection, we were concerned to include both process and effect oriented evaluations. Moreover, we wished to ensure a balanced distribution among the evaluations selected with 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.

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

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-Land 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.

The evaluators conducted a written survey in order to provide answers to the questions of interest (e.g. how strong was the perception of the campaign? Who accepts the contents of the campaign? How strong was the effect of the campaign?).

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. XX(4): 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 included different sources of information e.g. interviews, analyses reports and observations.

5) European Monitoring Centre on Racism and Xenophobia (EUMC)

Centre for Strategy & Evaluation Services (2002): *Evaluation of the European Monitoring Centre on Racism and Xenophobia. Final report*.
http://europa.eu.int/comm/employment_social/fundamental_rights/pdf/arcg/eumc_eval2002_en.pdf
(11.7.2005)

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 their study on EUMC publications, but also a survey and interviews were conducted.

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 BerufschülerInnen*. Bern: Edition Soziothek.

This project was 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 were evaluated using three written questionnaires.

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 theatre 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*. Journal of the American Medical Association. 277: 1605-11.

The violence prevention 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 behavior.

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 towards 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 program 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 project intends to educate adolescents during their vocational training through seminars and workshops providing them with knowledge about decision-making and responsibility when they encounter aggressiveness, violence and right-wing extremism.

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 audiences being reached are students and their teachers. The survey instrument chosen was a written interview. On the one hand, the study evaluated the effectiveness and thematic relevance for student purposes. On the other hand, it evaluated the impact of influence, pedagogic preparation and post-processing as an assessment of the project idea that could be used for teacher purposes.

12) CAT (Creative and Active Training)

GESOMED – Gesellschaft für sozialwissenschaftliche Forschung in der Medizin mbH (2003): *CAT – Creative and Active Training*. – not published
<http://www.powerforpeace.de/documents/cat-eval-2003.pdf> (11.7.2005).

This German project can be classified in the range of local networks. CAT is an educational intervention 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), USA. The seminars have been held in the following institutions: prisons, special school, secondary school and vocational school. In the evaluation, the process and effect of these trainings were taken into account. The evaluators developed therefore multiple course-specific questionnaires.

Results of the meta-evaluation

The quality of each of these studies has been thoroughly assessed with the SEVAL Standards (see section Approach and instruments above). Along the four standard groups of the SEVAL Standards the following results were achieved.

Utility

Overall, the standards criteria for this subject group were addressed to a satisfactory 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. Every evaluation clarified its objectives in a comprehensive way. In the majority of the cases, an absence of information about the 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 evaluate the competence 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 improvement are the descriptions of underlying approaches in order to assign value to the gained information. In general, the relevant information (e.g. program, context, aims, procedures and results of the evaluation) was available and could be understood easily. No evaluation informed about time frames. In some cases, the time period between the end of the project and the completion of the evaluation study was extremely short, so that we can assume that the evaluation provided the findings in a timely fashion. 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. The ‘Political Viability’ and the ‘Cost Effectiveness’ could not be judged because the corresponding information was missing.

Propriety

In the majority of cases, we were not able to assess the criterion ‘Formal Written Agreement’ for a lack of information. 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 securing during the planning and execution of an evaluation, it could not be explicitly seen from the texts, though it could be assumed. Due to missing information the standards ‘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.

Accuracy

The object of evaluation has always been described, documented and identified in an at least adequate way. Sometimes in short evaluations this description was not very detailed, but the evaluators made cross-references to further information sources. All evaluation studies share a well-done characterization of samples and key stakeholders. Differences arose in including the political or social context. In general the goals and questions pursued were communicated in a clear and comprehensive language. As goals, questions and procedures were not described sufficiently precisely, the evaluation process 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, one can say that the latter could not be judged as 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.). However, 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. 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 to the 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.

Summary

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 judgments 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 shows the results produced by the meta-evaluation.

Table 5: Results of the meta-evaluation

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 and 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 addresses the requirements imposed by the SEVAL Standards. In the case a judgment 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 and 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 opportunities for improvement of evaluation practice. No standard from the Propriety group is included here, 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.

Conclusion

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 deal directly with measures against right-wing extremism. Most of the others have measures against racism, discrimination, and violence as 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 an understanding of 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 measures could be identified for Austria and 12 for France, not a single project from these countries has reported an evaluation study to us.

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, some extensive, very well-documented studies exist 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 did not provide enough information about the methods applied. In addition, we have chosen a selection of studies representing the large diversity in terms of evaluandum and evaluation approach.

The meta-evaluation has shown that the quality of the studies under investigation is quite heterogeneous. 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.

Do the evaluation standards allow for a well founded assessment of an evaluation? Some interesting conclusions can be drawn from the meta-evaluations. The SEVAL Standards have proved to be an applicable instrument for the assessment of evaluation quality within the framework of a meta-evaluation. We strongly suggest that the SEVAL Standards are a useful tool to assess the quality of an evaluation, if meta-evaluation is conducted with a certain level of expertise in evaluation approaches, methods and techniques. Naturally, a thorough understanding of the standards is indispensable.

As the application of the standards has demonstrated, the standards are comprehensive, in the sense that they cover all relevant topics to consider evaluating evaluations. In addition, there is no indication that the standards in general or a specific standard is not applicable in the field of measures against right-wing extremism. The standards are methodologically open what is important because the evaluations of measures against right-wing extremism show a large diversity in epistemological terms.

The most salient problem we were confronted with during the meta-evaluation was clearly the lack of information. Since we based our investigations on the final evaluation reports only, we were not well enough informed to give an assessment for each standard. Especially with respect to the evaluation process and its context, there was a clear information deficit. This holds true – although to a lesser extent – as well for evaluation studies with an excellent reporting. Therefore, we recommend considering complementary information beyond the final report to reach a more complete assessment on evaluation quality in meta-evaluations.

Based on the empirical findings achieved in the meta-evaluation, we will continue in our research by formulating practice oriented guidelines how measures against right-wing extremism can be best evaluated. The quality aspects insufficiently addressed in current practice according to our meta-evaluation will be particularly emphasized in these guidelines in order to improve overall quality of evaluation in the field of measures against right-wing extremism.

References

- Ahlheim, K. (2003) 'Das Umfeld des Rechtsextremismus. Fremdenfeindlichkeit in der Mitte der Gesellschaft', in K. Ahlheim (Hrsg.) *Intervenieren, nicht resignieren – Rechtsextremismus als Herausforderung für Bildung und Erziehung*, pp. 25-48. Schwalbach: Wochenschau Verlag.
- Chen, H.-T. (1990) *Theory-driven Evaluations*. Newbury Park, CA: Sage.
- Falter, J. (1994) *Hitlers Wähler*. München: Westdeutscher Verlag.
- Furubo, J.-E., Rist, R.C. and R. Sandahl, 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, A. (2000) *Evaluation von Nachhaltigkeitsaspekten – Nachhaltige Evaluation?* Diploma Thesis: University of Vienna.
- Läubli Loud, M. M. (2004) 'Setting Standards and Providing Guidelines: The Means Toward What End?' *Evaluation* 10(2): 237-245.
- Leeuw, F.L. (2004) 'Evaluation in Europe', in R. Stockmann (ed.) *Evaluationsforschung*, pp. 61-81, 2. Auflage. Opladen: Leske + Budrich.
- Lehmann, L. and A. Balthasar (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 presented at the European Evaluation Society conference, Berlin, Germany: September 30 -October 2.
- Lulofs, K. and M. Arentsen (2001) *Improving Quality and Learning Performance of 'Energie 2000'*. Bern: Bundesamt für Energie.
- Nideröst, B. (2002) 'Erfolgsbedingungen für Evaluationen', *LeGes – Gesetzgebung & Evaluation* 13(1): 39-55.
- Pingel, A. und P. Rieker (2003) 'Evaluation pädagogischer Praxis gegen Rechtsextremismus: Potenziale und Chancen', in Von Berg, H.L. and R. Roth (Hrsg.) *Massnahmen und Programme gegen Rechtsextremismus wissenschaftlich begleitet*, pp.103-118. Opladen: Leske + Budrich.
- Rist, R. C. (ed.) (1990) *Program Evaluation and the Management of Government*. New Brunswick: Transaction.
- Scriven, M. (1991) *Evaluation Thesaurus*. Newbury Park, CA: Sage.
- Steffens, G. (2003) 'Rechtsextremismus als Extremismus der Mitte und die Aufgaben der politischen Bildung in der Schule', in E. Jung (Hrsg.) *Rechtsextremismus als gesellschaftliches Problem*, pp. 89-97. Schwalbach: Wochenschau Verlag.
- Widmer, T. (2005a) 'Instruments and Procedures for Assuring Evaluation Quality. A Swiss Perspective', in Schwartz, R. and J. Mayne (eds.) *Quality Matters*, pp. 41-68. New Brunswick: Transaction.
- Widmer, T. (2005b) 'The Development and Status of Evaluation Standards in Europe', *New Directions for Evaluation* 104: 31-42.
- Wüest-Rudin, D. (2002) 'Evaluation von Reformen der öffentlichen Verwaltung: Bedeutung und Erfolgsfaktoren', *LeGes – Gesetzgebung & Evaluation* 13(1): 57-82.