

Evaluators Evaluating Evaluators: - Peer-Assessments and Training Opportunities in Switzerland

MANFRED MAX BERGMAN¹

Abstract

This study examined how evaluators assess training opportunities and evaluations in their specific areas of expertise in Switzerland. Empirical data was collected from 15 exploratory interviews, 154 questionnaires, and 7 follow-up interviews. It was found that most evaluators have little formal training in evaluation; instead, they tend to rely on self-study, on-the-job training, and experience. Concurrently, most evaluators either do not know any evaluation training courses or do not know courses that are recommendable. The most recurrent complaint referring to training programs, as well as to evaluations currently executed in Switzerland, is the insufficient or inappropriate use of methodology. Other frequently mentioned shortcomings include the lack of cooperation either between evaluators themselves, or between evaluators and their clients. Various cost-effective measures are discussed that may alleviate some of these reproaches.

Résumé

Cette étude examine comment les évaluateurs jugent la qualité des formations et des évaluations dans leurs domaines d'expertise respectifs en Suisse. Des données empiriques ont été collectées, sous la forme de 15

¹ In collaboration with SANDRO CATTACIN, Laboratoire de Recherches Sociales et Politiques Appliquées, Geneva and MARLÈNE M. LÄUBLI-LOUD, Bundesamt für Gesundheit, Bern.

entretiens exploratoires, de 154 questionnaires et de 7 entretiens en profondeur subséquents. L'étude constate que la majorité des évaluateurs n'ont qu'une formation formelle réduite en évaluation; en contrepartie, ils se forment de manière auto-didactique, ainsi qu'au travers de leur expérience professionnelle. Il s'avère que la plupart des évaluateurs n'ont connaissance d'aucun cours de formation en évaluation, ou n'en connaissent aucun qu'ils recommanderaient. Le grief le plus fréquent à l'égard des programmes de formation actuels et des évaluations faites en Suisse concerne l'utilisation insuffisante ou inadéquate d'outils méthodologiques. D'autres lacunes fréquemment mentionnées sont le manque de collaboration entre les évaluateurs, et entre les évaluateurs et leurs clients. Cet article discute plusieurs mesures peu coûteuses qui permettraient de répondre à certains de ces reproches.

Zusammenfassung

Wie beurteilen Evaluatorinnen und Evaluatoren die Qualität der Ausbildungen und der Evaluationen, die in ihrem Bereich in der Schweiz durchgeführt werden? Um diese Frage zu beantworten, wurden Daten aus 15 Erkundungsgesprächen, 154 Fragebogen und 7 Tiefeninterviews gewonnen. Die Studie kommt zum Schluss, dass die Mehrzahl der Evaluatorinnen und Evaluatoren nur über eine rudimentäre und eher formale Ausbildung in Evaluation verfügen. Sie bilden sich vor allem autodidaktisch weiter und nutzen ihre Berufserfahrung. Ebenfalls mehr als die Hälfte der Befragten kennt gar keine Ausbildungsangebote in Evaluation oder kann bestehende Angebote nicht weiterempfehlen. Die Hauptkritik gegenüber bestehenden Ausbildungsangeboten und durchgeführten Evaluationen betrifft das mangelnde Methodenbewusstsein. Häufig wird auch bemerkt, dass es an der Zusammenarbeit mangelt, und zwar unter den Evaluatorinnen und Evaluatoren selber wie auch mit der Kundschaft. Der vorliegende Beitrag diskutiert verschiedene kostengünstige Massnahmen, mit denen einige der Mängel behoben werden könnten.

1. Peer-Assessments and Training Opportunities in Switzerland²

In 1987, the Federal Council of Switzerland launched a research program to study the effectiveness of various state regulated and state funded programs (PNR/NFP 27). According to the president of the expert group of this project, Ernst Buschor, this action was motivated by profound doubts about the efficacy of these programs (Buschor, 1997). Based on his findings, Buschor noted, among other things, the following problems with reference to policy evaluations and their effectiveness: first, political actors tend to dislike objective and external evaluation of their offices or programs, which may lead them to withhold information and otherwise obstruct the work of evaluators. Second, effectiveness is occasionally hampered by the inability of some actors or organisations to adjust rapidly or even at all. Finally, limited funds for evaluation or funds that tend to be partially controlled by non-evaluators, may constrain the quality of evaluations. According to Buschor, these factors may lead to incomplete or inadequate evaluations. Thus, he blames restricted access and insufficient funding as the main culprits for deficiencies in evaluations in Switzerland.

The Swiss Federal Office of Public Health (FOPH) believes that there may be another way of viewing the problem; that program evaluations may suffer from lack of expertise on the part of the evaluators. The FOPH is one of the major evaluation contractors in the public sector in Switzerland today. Indeed its policy development and implementation is increasingly based on evidence arising from the combined efforts of research and evaluation. One of its principal directives is therefore to monitor and improve its work by assessing the impact of its actions on the health of the nation's population (FOPH Directives, 1992, Bern).

² This project was funded by the Swiss Federal Office of Public Health (Contract 316.97.6773). We would like to thank Erik Verkooyen for his technical assistance with this project. Our thanks are also due to all the evaluators and their clients who have contributed so selflessly with their ideas and their time, and especially to those who were willing to be extensively interviewed. Commitment to confidentiality unfortunately does not permit me to name these individuals in person.

The FOPH established a specialist evaluation service in 1993 for co-ordinating and mandating its evaluation contracts. An assessment of training provisions and needs was identified as part of this service's efforts to develop quality standards for FOPH evaluation mandates. This research project is part of this effort.

This study primarily addresses the following questions:

- Have evaluators of Switzerland sufficient training and experience to perform state-of-the-art evaluations?
- Do current training programs in Switzerland offer sufficient and relevant training for professional evaluators?
- Are evaluators aware of such training programs?

According to our research question, we are not focusing on the interplay between clients, contractors, and evaluators, but on evaluators themselves. However, in no way are we opposing Buschor and his team's findings by blaming evaluators for possible shortcomings. Instead, we would like to point out additional obstacles that evaluators may face, i.e. insufficient training opportunities and deficient information diffusion of available training courses.

Given the difficulties we met in our initial attempts to locate the full range of training opportunities presently available in Switzerland and the reticence of training organisers to supply us with comprehensive details of their courses, we were unable to balance our assessment of the quality and scope of current training provision. As a result, our analysis is largely based on evaluators' evaluation of their evaluation training and praxis.

2. Evaluations and Professional Evaluators

What are evaluations, and why are they important? Areas of evaluation include, but are not limited to, general public health, substance abuse and mental health, education, business and industry, mass media, crime and

justice, management and supervision, special needs, feminist and minority issues, human services, international and cross cultural issues, needs assessment, product evaluation, state and local government, technology and research, and even theory of evaluation and evaluation procedures. Obviously, the range of programs, projects, or products that can be evaluated is extensive and, consequently, a definition of the concept, if it attempts to embrace most of the activities that fall under the rubric "evaluation," is likely to be vague. Hence, when Marlène Läubli-Loud (1997: 12) defines an evaluation as "the systematic collection and analysis of information not necessarily routinely available, about a specific project or programme to enable its critical appraisal," she must concurrently explain what a systematic collection and a systematic analysis are, what routinely available information is, and what distinguishes a specific project or a specific program from a non-specific project or program. Finally, she might have to explain what a critical appraisal is (compared to a non-critical appraisal). Ulrich Klöti's (1997: 39; cf. Bussmann, 1995) suggestions are similarly vague. He suggests that evaluations are "Informationsinstrumente ... [zur] Untersuchung der Wirksamkeit von Massnahmen und Programmen ... welche die Ermittlung der Wirkungen unterstützen können." To give further insight into the definition, the author has now to deconstruct his definition by explaining what instruments are, how efficacy is to be understood, and what constitutes an investigation into the effect of a measure or program. All authors, incidentally, elaborate on these issues in their writings. We merely emphasise here that definitions of the construct "evaluation" are by nature ambiguous, since they need to account for a tremendous variety of theoretical and methodological approaches and practices.

Evaluations are more important than ever because there hardly exists an aspect in either the public or the private sphere today which remains untouched by political, civic, or social programs or measures. Decisions taken based on the evaluators' reports may not only impact those people or programs under investigation, but also people or programs removed from the primary focus of investigation, but who interact with aspects under investigation. For instance, changes in governmental support for treatment programs for drug users will not only impinge upon the users themselves, but such policy changes will also have an effect on their activities and relationships with family members or close friends, and, in a wider sense, on how drug users will be framed and dealt with within

the larger society (e.g. criminals versus victims). Minor improvements, deteriorations, or just changes of a program or measure may mean gains or losses in terms of money, human resources, time, public support, etc. in related domains that can fundamentally impact the character of a company, health program, or public policy measure. The omnipresent impact of regulation of public and private spheres on the one hand, and the immense costs associated with such programs and measures on the other, no longer permit any government or institution to neglect aspects of efficiency, effectiveness, and cost-benefit analyses. This applies equally to an evaluation of a localised problem, i.e. work-shift regulation of a small company, or to a global evaluation of the regulation of the labour market on a supranational level.

3. Methodology and Approach

We targeted professional evaluators of Switzerland through interviews and via questionnaires. Access to mailing lists of SEVAL (Swiss Evaluation Society), FOPH (Federal Office of Public Health – Switzerland), and other contact lists formed the basis of our subject pool. We attempted to increase our pool further by contacting participants of past conferences on evaluation and by encouraging the questionnaire respondents of this study to reproduce the questionnaire and make it available to their collaborators. Based on this subject pool, we believe to have accessed most professional evaluators in Switzerland. To increase the response rate, 300 questionnaires were mailed in two waves, so each potential respondent received two copies of the questionnaire within one month. Eventually, we collected 154 questionnaires.³

Three sources provided us with the raw data for this study. First, we conducted 15 interviews that delimited the research domain in terms of evaluators' training experiences as well as their knowledge of existing

³ Because it is difficult to assess the size of our population since no comprehensive directory exists and since the evaluation activity itself is so difficult to define, we cannot guarantee that our sampling frame corresponds to the population. Nevertheless, responses from 154 evaluators of Switzerland make us rather confident that our results apply to the greater part of Swiss evaluators today.

training programs. Second, and in part based on the interview data, we constructed a questionnaire that allowed us to study especially these two aspects among Swiss evaluators in general. Finally, we conducted seven follow-up interviews with some specialists on evaluation, who assisted us further in interpreting our findings.

4. Results and Discussion

4.1 Some general characteristics of Swiss evaluators

Of the 154 responses received, men outnumber women by a ratio of 4 to 1. A gender analysis of the population based on the first name of the evaluators as it appears on our mailing lists revealed the same relationship. Thus, the ratio of one woman to four men cannot be explained by response tendencies. When asked why women are so underrepresented, one male interview respondent, a lecturer at a Swiss university, stated that this is to be expected in high-power positions. Asked why this should be expected, he claimed that a finding like this is typical and "repräsentativ für alle administrativen und amtlichen Positionen."

Most of the female evaluators in our sample (80%) were equally distributed across the age category 26 to 35 and 36 to 45. Only seven women were between 46 and 55. Most male evaluators, in contrast, were equally distributed across the age categories 36 to 45 and 46 to 55 (64%) – one full age category above that of women. 25% of the male sample were below the age of 36. The average age of men was more than 5 years higher than that of women.⁴

On average, our sample showed 7 years of job experience as active evaluators. However, there was a statistically significant difference across gender: while male evaluators have been active on average for

⁴ Mean difference: .58 of one age class which represents 10 years, i.e. somewhat more than 5 years. Independent samples t-test: $t=3.27$, $p<.01$ (adjusted for unequal sample size and unequal standard deviations).

about 8 years, female evaluators have been active for only 4.5 years.⁵ Male evaluators are therefore roughly 5 years older and have about 2 to 5 years more experience than female evaluators.⁶

4.2 Formal training of Swiss evaluators

Few of the evaluators of our sample had undergone significant formal training as evaluators. Many indicated short-term seminars (i.e. one to three days in length) or conference attendance which, for the purposes of this study, do not signify substantial training in evaluation. Others claimed self-training, on-the-job training, or professional experience. Finally, a number of respondents listed university degrees that seemed to be only secondary to their role as an evaluator. For example, they listed degrees in psychology, sociology, philosophy, mathematics, or physics in response to the question of which courses had facilitated their work as evaluators.⁷ Although undoubtedly valuable in their praxis as evaluators, we did not count these as evaluation-specific training. Of our sample, 56% listed either no substantial and formal training in evaluation, or training that was not directly related to that of an evaluator according to the criteria just stated.

Of those who reported some form of training (44%), 27% had received training of one year or less that directly pertained to performing evaluations, while only 17% had received training as evaluators for more than one year. In terms of university degrees that strongly implied training in evaluation (e.g. public health, public administration, public policy), our sample showed that 16 individuals held a PhD, 14 an MA or diploma, and 13 a BA or a license. This means that, according to the responses and

⁵ Independent samples t-test: $t=3.74$, $p<.01$ (adjusted for unequal sample size and unequal standard deviations).

⁶ 95% confidence interval based on unequal sample size and unequal standard deviations.

⁷ It certainly may be the case that sociology and psychology degrees, for instance, may include a course in public policy analysis. However, such courses are the exception to the rule. Our purpose here, as in all quantitative approaches, is to show rough tendencies; here, we will not attempt to describe in detail individual cases that fall outside of the general trends.

our criteria, about 28% of our sample held degrees directly related to program evaluation as defined above.

We do not claim here that those holding a degree in public health, for instance, are more qualified as evaluators than psychologists, sociologists, or mathematicians, who have never attended a seminar on evaluation. Similarly, we are not saying that a one-semester course in public policy prepares the evaluator better than a focused 2-day workshop. It is practically impossible to measure quality of training in terms of institutional affiliation or time spent in a classroom, particularly since evaluations vary enormously in terms of their nature and the desired goals of the evaluation. It is equally impossible to assess whether on-the-job experience for 20 years counterbalances a degree in public administration. Perhaps the public administration graduate is now miscalculating programs due to a lack of practical experience, but perhaps the senior evaluator without formal training has been miscalculating projects for the past 20 years. Consequently, these frequencies should be interpreted with great care.

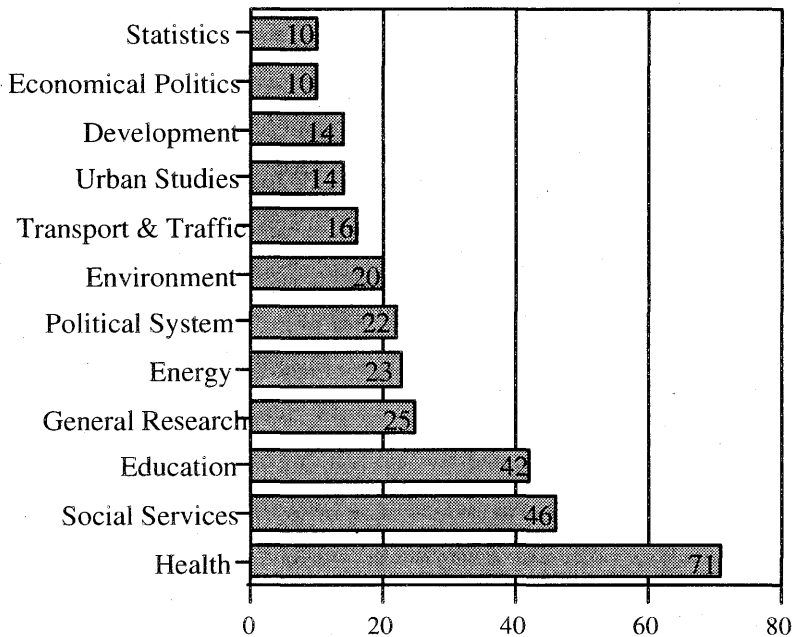
4.3 In which areas and in what positions are Swiss evaluators active?

The following table illustrates the main evaluation domains in which the respondents are active (table 1):⁸ Most of the evaluators in our sample are active in the domains of health, social services, and education. Less represented are the domains of research, energy, the political system, the environment, transport and traffic, urban studies, development and cooperation, political economy, and statistics. Evaluations in foreign politics, telecommunication, finance, justice, culture and media, agriculture, national security, and housing, were mentioned less than 10 times by our sample and therefore did not enter the graph below.

⁸ The total responses do not equal the sample size, because more than one answer was permitted. This also makes it impossible to analyse various response patterns according to evaluation area.

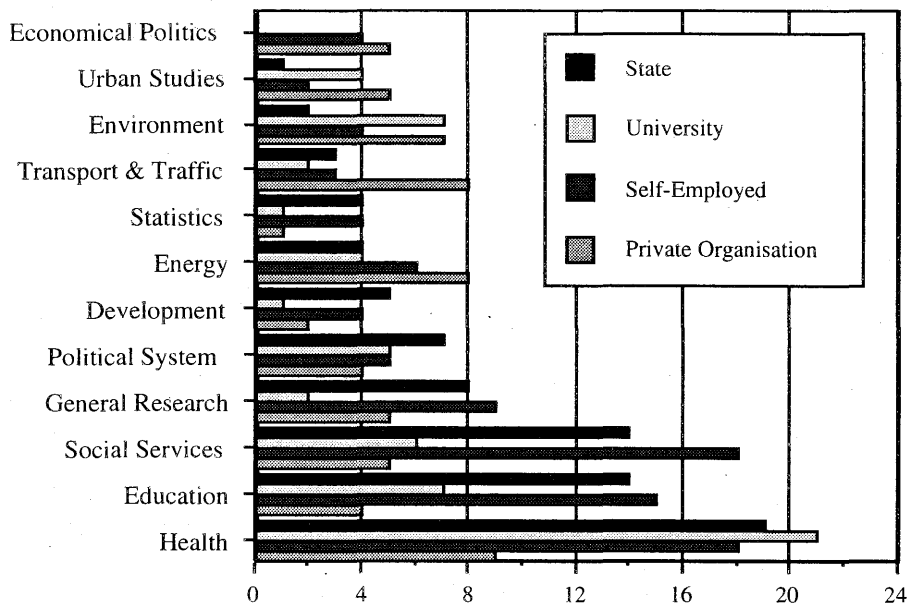
Table 1: Distribution of Swiss evaluators (n) by area of evaluation

The most common employer for the sampled evaluators was the state



(39%), while self-employment (23%) and the universities (20%) were mentioned with about equal frequency, and private enterprises were mentioned the least often (13%). Of the state-employed evaluators, most are mostly active in the domains of health, social services, and education. Evaluators employed by the university are active in the domains of health, education, the environment, and social services. Self-employed evaluators are mostly active in the domains of health, social services, or education, while evaluators employed by private enterprises are mostly active in the domains health, energy, transportation, and the environment. The following bar chart illustrates further the relationship between employer and evaluation domain.

Table 2: Distribution of Swiss evaluators (n) by area of evaluation by employer's domain



If there were no relationship between employer and evaluation domain, then the length of the bars in the chart representing the state-employed should be proportional to the amount of representation according to employer (i.e. 40% state-employed, 23% self-employed, 20% university employed, and 13% employed by private organisations or institutes). We can observe from this bar chart that the evaluation domains are not equally distributed across the evaluators' client. Health evaluations, for instance, are over-represented among university evaluators, while energy, transportation, the environment, and urban studies are over-represented among private institutions. In other words, although only 13% of the

evaluators in our sample are employed by private organisations, they nevertheless perform more than one third of the evaluations on the environment and almost half of those in urban studies and transportation. Concurrently, although 40% of the sampled evaluators are state-employed, only 27% are active in the health domain.

4.4 Evaluation designers versus evaluation leaders

There is an obvious association between designing and leading an evaluation project,⁹ but what are the best predictors for being an evaluation designer or a leader of an evaluation project? Being the designer of an evaluation project was best predicted by years of professional evaluation experience and not by formal education as an evaluator, age, or gender.¹⁰ However, the variables predicting leadership in an evaluation project were years of experience and, controlling for this effect, also gender.¹¹ Years of practice as an evaluator is therefore the most important predictor of whether an evaluator is or is not the project designer and leader. While there is no effect of gender on the likelihood of designing a project, being a woman results in a decreased likelihood of also being the project leader. This is particularly interesting, since women have on average about 3 years less experience in project evaluation, compared with men. It appears that this lesser experience is reflected not in their roles as project designers but in their roles as leaders.¹² Formal education conti-

⁹ Pearson's $r=.56$, $p<.001$.

¹⁰ Multiple regression: $R=.28$ ($F=2.84$, $p<.05$); standardized partial regression coefficients and significance levels: years as evaluator: $\beta=.30$ ($p<.01$); formal education as an evaluator, age, and gender: $p=n.s.$ Logarithmic transformation of the initially skewed kurtotic dependent variable did not change the results significantly.

¹¹ Multiple regression: $R=.40$ ($F=5.09$, $p<.01$); standardized partial regression coefficients and significance levels: years as evaluator: $\beta=.32$ ($p<.01$); gender: $\beta=.21$ ($p<.05$); formal education and age: $p=n.s.$ The results of this analysis did not change significantly, after the initially skewed and kurtotic dependent variable was transformed via the natural logarithm.

¹² This finding is substantiated by Pearson's χ^2 -Test of Independence (for designer and gender: $\chi^2=1.5$, $p=n.s.$; for leader and gender: $\chi^2=9.4$, $p<.05$) as well as the correlation coefficient of Pearson's r (for designer and gender: $r=.07$, $p=n.s.$; for leader and gender: $r=.25$; $p=.01$).

nues to play an interesting role because it predicts neither project design nor leadership. Because of the relatively low level of formal education in evaluation, we may be dealing with a floor effect, i.e. the variable "formal education" is not sufficiently distributed toward the higher end to register a relationship.

4.5 Subjective assessment of evaluation courses

By far the most frequently mentioned institution where the respondents had received evaluation-specific training, regardless of duration of training, was the University of Bern (n=27). Other Swiss institutions mentioned more than two times were, in alphabetical order, FOPH¹³, DEZA, ETH, IDHEAP, NADEL, SEVAL, the University of Geneva (especially DESMAP), the University of St. Gall, and the University of Zurich, but none of these were mentioned as frequently as the University of Bern. Twenty evaluators of our sample received substantial evaluation training in England, France, Germany, and North America.

But the courses which Swiss evaluators recommend, and qualities which they look for in their collaborators and employees, are a different matter. These issues are obviously dependent on the area of evaluation, the specific role that the evaluator is expected to fulfil (e.g. monitoring, evaluation design over a given time period, etc.), and abilities which the evaluator already possesses. Considering these limitations, we can nevertheless observe which institutions are most frequently mentioned.

Respondents were asked which training courses pertaining directly to evaluation they would recommend the most. Just over 70% of our sample either did not respond to this question at all, or stated that they did not know any evaluation courses in Switzerland, or that there were no recommendable evaluation courses. Of the 45 respondents who were able to make recommendations, most mentioned the University of Bern (both continuing education courses and regular university courses; n=17) and

¹³ The Swiss Federal Office of Public Health, itself does not actually run courses on evaluation. However, its evaluation unit has funded occasional short seminars, the majority of which have been organised by university institutes. This could therefore explain this confusion from the participant's point of view.

the IDHEAP (n=10). Other programs recommended between 3 and less than 10 times were, in alphabetical order, the FOPH, ETH, seminars organised by SEVAL, the University of Geneva (especially DESMAP), and the University of Zurich. We may conclude that Swiss evaluators on the whole are relatively poorly informed about available evaluation courses. If they know about them, they seem to be unable to make statements about their quality, or find that these courses are not recommendable.

We asked those who were familiar with some of the programs in Switzerland, which general problems they perceived in the currently available evaluation courses. The responses of those who were familiar with at least some courses were so multivariied and significant that we will elaborate on this point in greater detail. One recurring theme revolved around the superficiality of the courses. While some claimed that introductory courses to evaluation were too short and superficial – thus making them a "waste of time" – others found even more advanced courses unchallenging and irrelevant for their specific needs. One respondent complained: "Il y a une manque énorme de cours pour des évaluateurs avancés pour améliorer leur compétences. Il y a trop d'introduction à l'évaluation qui ne sert à rien." Overall, both beginners and advanced evaluators found the programs too superficial, and thus seemed to prefer self-directed studies. A related problem seems to be that the interests of participants in these courses and training seminars are too heterogeneous. In other words, many participants become frustrated by the courses' lack of specificity because courses attempt to fulfil too many and too varied needs.

A few respondents complained about the length of the courses or programs. Those referring to university courses suggested that they should be more concentrated and to the point, while those referring to short-term seminars suggested that these should be extended by a few days (e.g. from 2-day to 5-day intensive seminars). This would supposedly solve the "waste of time" problem expressed with reference to university courses, and also address complaints about the superficiality of short-term courses. One interviewee suggested that universities should open up their courses to professionals, i.e. that it should be possible to follow a university course without having to sign up for the entire degree program.

Finally, many complained that the training on both the university level and in short-term seminars was not sufficiently oriented toward the practical requirements of evaluators. It was mentioned repeatedly that university courses were too theoretical and not sufficiently praxis oriented, and that non-university courses were not rigorous enough in terms of theory and methodology. One respondent summed up the concerns of many: "Es fehlt an Transparenz um das Weiterbildungsangebot ... und es fehlt ein Ausbildungsgang, der methodologisch fundiert und gleichzeitig ausreichend praxisbezogen ist."

Although it seems that evaluators ideally wish to attend courses that teach all there is to know about evaluation theory, method, and praxis (plus the course should concentrate on their interest area only) – and all that preferably in less than one week, this should not lead course designers and teachers to throw up their hands in exasperation. Various steps could be undertaken to better serve the training needs of evaluators. For instance, some interviewees recommended the inclusion of an apprenticeship into a university degree program, which would familiarise the student with the practical aspects of evaluation. Evidently, the same recommendation cannot be made with short-term courses. None of the interviewees had suggestions on how these could improve their applicability in "real-life" evaluation situations. It might be possible, however, to leave some time at the end of such short-term seminars to perform individual "tutorials" or "clinics." In other words, why not give the participants in mini-seminars the opportunity to discuss their specific problems as they relate to the subject matter with the course convenor at the end of the course? On the one hand, this would not frustrate the other participants because it would avoid a diversion from the main topic during the seminar, while on the other hand, it would still give the participants the opportunity to connect the newly acquired material with their specific problems.

Overall, it can be stated that many evaluators are either uninformed and/or dissatisfied with the currently available evaluation programs on both university and non-university levels. SEVAL has been making corrective steps toward information diffusion, but this does not seem to be enough. An excellent and cost-effective alternative would be to exploit the internet. Evaluators and their organising bodies in some other countries make far better use of the internet than do the Swiss. For example,

there exist more or less extensive internet sites on evaluation for countries such as the US, Canada, England, and Germany (see the appendix for the various http addresses). A similar effort by Swiss evaluators could be made to better inform each other on relevant events and to improve the general exchange of information, methodology, and pertinent research results. Furthermore, such an internet site could be used by *clients*, not only for posting upcoming projects (which would serve both clients and evaluators), but also as a source of reference or information about the work of evaluators in general, and for specific events relating to evaluations in particular. Additionally, a sophisticated and well-managed site would permit publication of an on-line evaluation journal that would help to diffuse research findings in all areas of evaluation. Not only would this allow evaluators to publish articles and research reports in areas that are under-served by the current outlets, but it could easily develop into a major encyclopaedic resource. Given sufficient resources, it would be rather easy to use the internet to create a site dedicated to evaluators and their concerns. But the internet represents far more than a gigantic library on the computer screen: it could serve as a communication node where a political scientist, a health program evaluator, a legal advisor, and a specialist on primary education could find a forum for exchanging ideas – something that is unlikely at present.

For more practical and short-term solutions, it should be mentioned that there exist excellent discussion groups on the internet, where both beginners and advanced evaluators are able to post their queries (see the appendix for some discussion groups). Despite the variability of the quality in the responses from internet discussion groups, evaluators are nevertheless likely to obtain excellent advice for even the most specific and complex problems.

4.6 Preferred qualities of evaluators

The next aspect to be analysed is the demands which evaluators make on potential co-workers. In other words, respondents were asked which evaluation-specific qualities they would look for in a future collaborator or employee. A clear tendency emerged here: evaluators judged methodological knowledge (n=63) nearly twice as important as the second

highest qualities (n=36 for issue-specific knowledge and n=33 for prior experience as an evaluator). In conjunction with the findings from earlier questionnaire items, methodology keeps re-appearing as a very important, yet neglected aspect. It furthermore implies that a well-grounded and substantial familiarity with diverse methods is difficult to acquire in short-term seminars. Here again, it might be of advantage to open up semester-long methodology courses at the university level to evaluators or, even better, design or improve the design of courses on a university level that specifically address the methodological requirements of evaluators. Additionally, it is imperative that information on evaluation training courses at all levels is better diffused so as to permit evaluators who are not affiliated with a particular university or department to be aware of such opportunities.

Alternatively, courses offered by reputable summer schools tend to be as thorough as semester-long university courses - at a considerable gain in time. Acquiring as much knowledge within one to three weeks in the summer as can be expected from a semester at a university should be an attractive incentive for those who wish to acquire a solid methodological knowledge base. Moreover, some summer schools offer methodology courses for all levels and in various areas which further serve those with specific methodological needs. Excellent summer courses in research methodology and data analysis which are open to the general public are offered by the University of Essex and the University of Michigan. Another spring course is offered by the University of Mannheim, while yet another - apparently limited to one or two themes per year - is offered by the University of Lille (see the appendix for more information). Two years ago, Switzerland started its own annual summer school in data collection and analysis, but, at least for the present, it is limited to students enrolled in PhD programs.

Until there is a summer school that addresses the specific needs of evaluators, which may occasionally be different from the needs of social scientists in general, attendance in the above mentioned summer courses may help to reduce many of the shortcomings that were mentioned by the respondents.

5. Self- and peer-critique

The final response cluster to be analysed in this study revolves around criticisms which evaluators raised with respect to evaluations performed in their field. More than half of our sample provided significant and occasionally very detailed concerns, which deserve to be elaborated in detail. Overall, the responses centred around methodological concerns, interchange between evaluators, interchange with clients, quality control, and concerns about ethics.

5.1 On Methodology

Here again, methodological problems were raised most frequently. Two-thirds of the respondents mentioned various methodological aspects, ranging from "spongy" conceptualisations of the evaluation to sloppy reporting of results. There were recurring themes that are worth mentioning: methodological criticism frequently included the lack of structure or of explicitly stated research design, approaches, or goals; ambiguous or insufficient operationalisations; deficient sampling and data collection; too much *ad hoc* theorising; lack of methodological pluralism (including a lack of triangulation); exaggerated differentiation between qualitative and quantitative methods; unsophisticated or incorrect application and interpretation of statistics; unfounded causal statements; unrepresentative and overextended results; development of evaluation concepts and theory during the write-up period (i.e. the end of the study); reliance on circumstantial, trivial, and anecdotal evidence; and a lack of checklists and research protocols.

5.2 On Interchange between Evaluators

Although this section diverges somewhat from the theme of this paper, we think that it might still be useful to some readers. Evaluators frequently complained about the lack of interchange between evaluators on various levels. This included the exchange of experiences, research results, their final reports, and relevant publications, but also a lack of access to others' raw data. In this context, SEVAL was occasionally men-

tioned as an important organisation that has already served the evaluators of Switzerland, but some felt that SEVAL should expand its focus even more. First, respondents felt that the organisation is too narrowly focused and field-specific (i.e. political science and public administration). Second, SEVAL's efforts were praised in terms of representation and collective defence of evaluators' needs, but it was suggested that SEVAL should expand its function to contribute to networking in terms of information diffusion regarding open competition for evaluation projects and training seminars on methodology at all levels.

Given the potential importance of SEVAL as the overall co-ordinating body for evaluators' interests in Switzerland, various committee members at SEVAL were asked about these possibilities. The following should not be read as an official response of SEVAL but only as the private opinion of some relatively high-ranking members of the organisation. One subgroup suggested that SEVAL is not likely to expand such that other fields of evaluation would find a stage for their issues. They reasoned that the directorship of SEVAL currently consists of political scientists and administrative specialists. Even if SEVAL were to be able to dedicate some of its organisational resources to other areas, such as education, health, or feminist/minority issues, evaluators in these areas would nevertheless fail to be sufficiently served because, at present, there does not exist the knowledge base and structure to accommodate these substantially different fields of interest or research. One officer, however, expressed great interest in expanding SEVAL in the following ways: first, to integrate into the organisation evaluators who are not primarily interested in policy and public administration and, second, to play a more active role in network building. He mentioned the possibility to accept executive members whose primary interests do not lie in politics or administration. Third, additional staff positions, e.g. secretarial support, would make it possible not only to increase the information diffusion necessary for better networking, but it would also allow to increase SEVAL's membership by finding and contacting evaluators not yet integrated into the organisation.

Should SEVAL indeed be unable or unwilling to expand in order to accommodate the many other evaluation fields, then there seem to be only two possibilities left for those not currently represented. First, evaluators from underrepresented evaluation areas would need to collectively or-

ganise and create their own body of representation which would both defend their interests and take on the task of information diffusion and networking among the members. The FOPH, for instance, has recently established its own directory of evaluators experienced in the field of health. Currently this is essentially used for "calls for offers" for its evaluation studies, but it could be developed to cover other areas and facilitate networking in general. For other organisations and sub-fields, this would be exceedingly difficult because evaluation is not as well established in Switzerland as it is in some other countries. Additionally, since Switzerland has a rather limited number of evaluators, especially in areas other than public policy, administration, or health, it would be highly inefficient and possibly ineffective to create multiple organisations, consisting of only a few members which would have to largely replicate functions that seem to already work rather well at SEVAL.

A second possibility would be to create a parent-organisation of all evaluators of Switzerland, in which SEVAL would be one division among many. Understandably, this was strongly rejected by some of the SEVAL officers. As a well-functioning organisation, SEVAL would not be willing at this point to subordinate itself to a superordinate organisation. Apparently, SEVAL serves most of its members well and, accordingly, there is no need to make changes which might endanger the status quo.

Therefore, evaluators who remain marginal in terms of the representation and interest of SEVAL are – at least for the time being – left to their own devices. However, those who feel disconnected from the field of evaluation in Switzerland may find a welcoming public and a forum in other countries where associations of evaluators are indeed subdivided into various divisions (see the appendix for some resources).

5.3 On Interchange with Clients

Evaluator-client relations should be part of the training, although this dimension, according to some respondents, was lacking entirely in most training seminars. For instance, it was suggested that clients' expectations are often unrealistic in terms of what an evaluator may be able to accomplish. According to the interviewees, this leads to unkept promises

on the side of the evaluator and disappointments on the side of the client. Clients should have at least some nominal training which would permit them to have some rudimentary insight into the work of evaluators (see, for instance, Läubli-Loud, 1997 and Bussmann, 1995a, for introductory evaluation texts that could be useful also for clients).

For others, short-term courses may be able to teach *clients* how to avoid many conceptual errors in their mandates. This request parallels demands for more interaction between client and evaluator, which is likely to be most fruitful if clients have at least a working knowledge of the field of evaluation. Evaluator-client interactions backed by knowledge of each others' domain are likely to lead to a better appreciation of both the nature of the work and the research results. "Educated" clients are potentially more likely to accept unfavourable results, a problem which has been raised as well.

It may be feasible to diffuse information about, as well as to actually organise tuition-free one-day courses on, evaluations which centre specifically around the needs of clients.¹⁴ It is in the interest of evaluators to produce a short and informative pro-to-lay person pamphlet (e.g. of 10 pages) which could explain the main aspects of the work of evaluators. This source should furthermore include web-site information, evaluation-specific publications relevant for clients, and other contact addresses for those who wish to obtain further information. This pamphlet should be distributed - free of charge - to all key clients and organisations.

Beyond increasing clients' competence in writing coherent and realistic evaluation mandates, these education efforts might lead them to obtain instructions on how to evaluate evaluators' offers. Finally, such encounters are likely to increase the interaction between clients and evaluators, as well as the appreciation of the different problems with which *both* sides are confronted. Various respondents stated that evaluations must be integrated within virtually any larger project or program. Convincing the sometimes suspicious client about the advantages of monitoring and evaluation-as-process would be substantially facilitated by short-term

¹⁴ I am convinced that clients *but also evaluators* could learn a lot about the client's role in the evaluation context.

client-based information courses and the distribution of a condensed yet accessible pamphlet.

Beyond unrealistic expectations and errors in the conception and elaboration of the mandates, respondents also mentioned that open competitions for projects needed to be better publicised. An internet-based information centre dedicated to the evaluation community (where community is meant in its broadest sense, i.e. integrating clients as well) could also serve this purpose very well.

5.4 On Quality control

Numerous respondents raised the issue of licensing. They stated that advanced graduate courses, certificates, postgraduate degrees, and official accreditation would assist in the professionalization of the field of evaluation. Beyond status gains, moves in this direction might raise minimum standards and qualifications. Concurrently, however, it would also restrict access according to *a priori* rules, which are extremely difficult to defend, given the ambiguous and context-specific character of evaluations. As one of the interviewees pointed out, evaluations as causal and analytical undertakings have a substantially different character compared to evaluations in marketing and opinion research, which are different again from evaluations relating to ethics, justice, insurance, national safety, or national health. A conceptualisation of unifying standards which are nevertheless still sensitive to the field-specific character of these evaluation families seems exceedingly difficult. Others insisted that the field of evaluation must not professionalize but must be made a mandatory and integrated subject within any social science degree.

In response to quality concerns, two interviewees believed that "the market of evaluation" would take care of the underqualified evaluators. The market, according to the respondents, would eventually support the good evaluators and assure that the reputation of those who perform inadequately bars them from being employed as evaluations in the long run. This "Just-World-Hypothesis" seems to be a risky gamble, even if it would possibly lead to desired results. Because evaluations are relatively new in Switzerland, and because relations between clients and evaluators are still in need of improvement, such a stance may turn out to be coun-

terproductive to the goals of establishing evaluation as a matter-of-course activity on the one hand, and improving the overall quality of evaluations, on the other. How likely is it that someone will spend another half-million francs on this year's evaluation if the first half-million did not bring last year's promised results? And lastly, it is not at all certain that the market does take care of itself – especially if many clients do not have the tools to distinguish sufficiently between a good and a bad evaluation (as opposed to desirable and undesirable evaluation results).

While the group favouring regulation of the field of evaluation tends to seek official accreditation and recognition, the group favouring a de-regulated market resists professionalization by arguing that evaluation outside of a context (i.e. a subject which is to be evaluated) is bound to fail, because the discipline will become self-contained and self-serving. As briefly discussed, there are substantial pitfalls in either position, should they be defended too uncritically.

Interviewees also lamented the fact that universities are becoming market-driven evaluation organisations. They argued that universities are removing themselves from the sometimes unremunerated work of theory building and testing, but are instead trying to break into the lucrative private sector. With universities, private organisations, and independent evaluators all competing for substantial funds, so the interviewees argued, the university is abandoning a check-and-balance function. Rather, it is becoming just another player for coveted funds. It was thought to be preferable that universities remain independent, i.e. limit their competition with the outside market and maintain quality standards as well as a supervisory, didactic, and critical role which would be compromised by market-driven interests.

A related issue was raised in terms of unfair competition. Numerous interviewees complained that university-based but largely professional evaluators can use the significant infrastructure of the institution to promote their private interests. Various overhead costs arising from office rent, personnel, computing, consumable materials, etc. are covered by the university (i.e. tax money), while evaluators not linked to the university structure have to pay for these themselves. For this reason, the latter group is forced to cut corners with respect to quality in order to be able to compete against university-subsidised evaluators, who have a multi-

million francs strong organisation as well as eager, compliant, and relatively inexpensive graduate students at their disposal.

Many evaluators expressed concerns about the results sections of the final reports of evaluations. One sub-cluster emerged that targeted the lack of practicality, cost effectiveness, and feasibility of proposed solutions. One evaluator, who is also responsible for contracting other evaluators, stated that evaluators "should not be surprised if their findings 'disappear' into the drawer because their suggestions are often for the dogs." Yet another respondent proposed that the solution package should not contain ambiguous suggestions such as "changing the system at the highest level" without elaborating on such "fishy" notions. Another declared that evaluators must remove themselves from "Pseudodiskussionen, Elfenbeinturmkauderwelsch, und banalen und schwammigen Schlussfolgerungen."

Overall, many suggested that evaluators in Switzerland needed to bring more formal training, as well as more multidisciplinary approaches, to the task. In this context, some recommend continuing education and regular participation in conferences, regardless of professional accomplishment. This text will not treat this subject in more detail, but the interested reader is referred to Widmer (1996) who offers further suggestions.

5.5 On Ethics

Dependence on funds and on clients was mentioned as the prime source of ethical conflicts. Some respondents believed that their funding and extensions of a work contract depended not on the quality of the research but, at least in part, on the findings themselves. Although the respondents insisted that they themselves had never falsified results in order to continue a work relationship with their clients, they claimed that they knew of other evaluators whose desire to continue working for a client led them to opt for diplomacy rather than for an objective presentation of the research results. Additionally, evaluators are sometimes forced to make unrealistic promises in order to be hired by the client. Accordingly, this breeds not only disappointment with the end result on the side of the client, but also an unethical work attitude among evaluators. If the only

way to obtain a contract is to exaggerate possibilities and abilities, then evaluators are either forced to compromise their ethics or lose out to those who have no problem with making unfounded claims or overoptimistic promises. Here, clients need to learn – possibly during short-term seminars – that the proverb "You get what you pay for" often applies to their domain as well.

Various respondents raised other ethical issues, such as unfamiliarity with research and profession-specific ethics guidelines, laws of data protection, and how to diffuse vital information that the client does not want to be diffused.

It seems to be necessary to begin thinking about the elaboration of a code of ethics for evaluations in Switzerland. Although such a code is not likely to make unethical evaluators more ethical, it will clarify ethical positions for those who are unclear about them in the first place. And for those who decide to behave unethically, a code of ethics will at least make unambiguously clear that they are, indeed, acting unethically.

6. Conclusions and Recommendations

Bearing in mind the criticisms made by many evaluators with regard to research ethics, I will try to make some concrete suggestions without being overly worried about either diplomacy or the strong responses that these will undoubtedly draw.

The most significant finding of this study is that evaluators on the whole have little significant evaluation-specific training. Although it could be argued that this is counterbalanced by substantial on-the-job experience, our analysis of the self- and peer-group criticism of evaluators undermines this view. Especially with reference to methodology - a recurring theme - evaluators judge both courses and evaluations performed in Switzerland as deficient.

Second, we found that universities were evaluated as not praxis-oriented enough, and non-university training programs as too specialised, unchallenging, and non-theoretical.

Third, it was found that there is a lamentable lack of information diffusion on training opportunities as a whole. An extensive internet site – possibly including an on-line multidisciplinary evaluation journal – seems an *essential* first step in developing an evaluators' network as well as grounding evaluators, who frequently complained of "floating around."

Fourth, it is not clear whether evaluators in general are for or against a professionalization of their field. On the one hand, a professional status may assist in quality control and rigor, while on the other hand, some oppose this because they fear that evaluations will distance themselves from their subject. Nevertheless, advanced courses or even a postgraduate degree in evaluation would certainly move in the direction of responding to the recurrent demand for higher overall standards. To avoid criticism by those who fear a self-containment and subject-distance, such a degree must somehow continue to be founded on specific areas of evaluation, possibly as a type of specialisation that is directly connected with the degree in evaluation (e.g. diploma in evaluation with a specialisation in drug policy; MA in evaluation with a specialisation in public health policy).

Fifth, client education appears to be very important because it may resolve not only the ambiguity and mistrust that seems to exist between some clients and their evaluators, but may also modify the unrealistic or ambiguous evaluation expectations, as well as the rather serious ethical problems arising from unpopular research findings. Diffusing information on evaluation as an activity, as well as organising short but informative courses on what evaluators can do and how to pick the right evaluator for the program or project at hand should take care of the more basic needs in this domain.

Sixth, most evaluators displayed a surprising lack of insight and empathy with regard to their clients' problems. Throughout formal interviews and informal discussions with evaluators, it appeared that clients were often considered short-sighted, stingy, and ignorant about the problems evaluators have to face. However, it appears to me that most evaluators are not very knowledgeable about problems that *clients* undoubtedly have to struggle with. Neither do clients have unlimited funds at their disposal, nor are they completely free to negotiate evaluation conditions - contrary

to the beliefs of some evaluators. As a follow-up to this project, it would be interesting to devise a study - paralleling the methods and focus of this project - in which clients discuss (a) their own problems with regard to hiring and assessing evaluators and their products, (b) their training experience and needs with regard to evaluations, and (c) their suggestions on how to improve evaluations, as well as relations with evaluators.

All of the criticisms raised in this paper are based on the individual evaluators' perception of their field and should therefore not be misinterpreted as either a collective comment or as necessarily reflecting the opinions of the author. In other words, I am neither defending all the criticisms raised here, nor am I claiming that the occasionally rather severe criticisms voiced by the respondents represent accurate assessments of the field. But especially due to the lack of information diffusion and a rather limited experience with formal courses, evaluators themselves are critical toward their field and their colleagues. It is impossible to satisfy the diverse and sometimes incommensurable needs of the individual evaluators or clients, but that does not mean that needs cannot be satisfied at all.

On a more positive note, we believe that Switzerland offers excellent opportunities for capable evaluators, as well as for those who want to become capable evaluators. The field of evaluation is rather new and therefore still malleable. With good intentions and by following them up with substantial, appropriate, and far-sighted deeds, it is possible to change the overall quality as well as the image of the entire field of evaluations for the better.

7. References

- BUSCHOR, E., Vorwort. In: W. Bussmann, U. Klöti, & P. Knoepfel, *Einführung in die Politikevaluation*. Basel 1997.
- BUSSMANN, W., Instrumente der Erfolgskontrolle, *VOP (1995)*, S. 345-351.
- BUSSMANN, W., *Evaluationen staatlicher Massnahmen erfolgreich begleiten und nutzen: ein Leitfaden*. Chur/Zürich 1995.

- BUSSMANN, W., KLÖTI, U., & KNOEPFEL, P. (Eds.), *Einführung in die Politikevaluation*, Basel 1997.
- KLÖTI, U., Inhaltliche und methodische Anforderungen an wissenschaftliche Politikevaluationen, In: W. Bussmann, U. Klöti, & P. Knoepfel (Eds.), *Einführung in die Politikevaluation*, Basel 1997.
- LÄUBLI-LOUD, M., *Guidelines for health programme and project evaluation planning*, Bern 1997.
- OFSP, *Mesures fédérales pour réduire les problèmes de la drogue*, Bern 1991.
- WIDMER, T., *Meta-Analysis: Kriterien zur Bewertung von Evaluationen*, Bern 1996.

Appendix

1) Some Selected Internet Sites on Evaluation

- <http://www.admin.ch/bj/rspm/rspm-d.htm> (judiciary issues and evaluations)
- <http://www.admin.ch/bj/rspm/evkurse/agenda-d.htm> (continuing education for evaluators in Switzerland)
- <http://www.europeanevaluation.org/> (European Evaluation Society)
- <http://www.oecd.org/puma/country/switzerld.htm> (Public Management Service, i.e. PUMA)
- <http://www.evaluation.org.uk/> (UK Evaluation Society)
- <http://www.eval.org/> (American Evaluation Association)
- <http://www.eval.org/ElectronicLists/evalinfo.html> (distribution of information of evaluation related materials)
- <http://www.fal.de/~tissen/geproval.htm> (Deutsche Gesellschaft für Evaluationen)
- <http://www.valutazione.it/> (Associazione Italiana di Valutazione)
- <http://www.unites.uqam.ca/sce/ces-sce.html> (Canadian Evaluation Society)
- <http://203.32.109.1/aes/> (Australasian Evaluation Society)

2) Discussion Groups on Evaluation on the Internet

<http://www.eval.org/ElectronicLists/evaltalk.html> (general discussion group for evaluators)

<http://www.unites.uqam.ca/sce/links.html#GOVTEVAL> (discussion group for public sector evaluation)

<http://www.unites.uqam.ca/sce/links.html#XCeval> (discussion group for cross-cultural evaluation)

3) Some Internet Sites for Summer Schools in Data Collection and Analysis

<http://www.essex.ac.uk/summer98/> (University of Essex Summer School of Data Collection and Analysis)

http://www.icpsr.umich.edu/ICPSR/Other_Resources/Summer/summer.html (University of Michigan Summer Program in Quantitative Methods)

Other summer courses are advertised within the specific sites listed in **1**).

