Ilse Julkunen (Ed.)

Perspectives, models and methods in evaluating the welfare sector - a Nordic approach

The publication can be ordered free of charge from the address:
STAKES
National Research and Development Centre for Welfare and Health
Evaluation of Social Services (FinSoc)/Hilkka Nyström
P.O.Box 220, FIN-00531 Helsinki, FINLAND

Phone: +358 9 39671 and 3967 2313
Telefax: +358 9 3967 2278
Email: hilkka.nystrom@stakes.fi

Johdanto

Ilse Julkunen

Huhtikuussa 2003 alkoi Helsingissä seminaarisarja Arviointi Pohjoismaissa – pohjoismaisia malleja ja menetelmiä. Sarjan suunnitteli FinSoc yhdessä Helsingin yliopiston Svenska social- och kommunalhögskolanin ja sosiaalipolitiikan laitoksen sekä ruotsinkielisen sosiaalialan osaamiskeskuksen kanssa. Sen tavoitteena oli tutustua pohjoismaisiin kokemuksiin arvioinnista ja erilaisista arviointimalleista sosiaalialalla.

Seminaarisarja koostui kolmesta kahden päivän seminaarista, joissa sekä pohjoismaiset että suomalaiset sosiaalialan tutkijat ja ammattilaiset luennoivat. Ensimmäinen seminaari pidettiin 14.–15.4. 2003, ja siinä käsiteltiin arvioinnin kehitystä ja merkitystä sosiaalialalla ja konkreettisemmin paneuduttiin arviointimalliin, jossa arviointia käytetään osana omaa työtä. Seminaarin vierailevat luennoitsijat olivat Evert Vedung Uppsalan yliopistosta ja Göran Sandell Göteborgin yliopistosta. Toinen seminaari pidettiin 15.–16.9.2003, ja siinä vierailevana tutkijana oli Hanne Kathrine Krogstrup Aalborgin yliopistosta. Tässä seminaarissa keskityttiin siihen, miten asiakkaat voivat osallistua arviointiin ja miten asiakkaiden mielipiteet saadaan kuuluville erilaisten arviointimallien avulla. Lisäksi käsiteltiin arviointia kehittämisvälineenä. Viimeisessä seminaarissa 10.–11.11.2003 keskusteltiin siitä, miten vaikuttavuutta voidaan arvioida sosiaalityössä. Lars Oscarsson Örebron yliopistosta oli vieraileva luennoitsija. Yksityiskohtaiset ohjelmat sekä koulutusmateriaalit löytyvät FinSocin kotisivulta www.stakes.fi/finsoc.

Seminaarikielet olivat ruotsi ja englanti, mutta keskustelua käytiin myös suomeksi. Seminaaripaikka oli Svenska social- och kommunalhögskolan, ja jokaisessa seminaarissa oli noin 80 osallistujaa. Osa osallistujista oli mukana kymmenen opintoviikon arviointikurssilla, jonka Helsingin yliopiston jatkokoulutuskeskus Palmenia järjesti.

Tämä raportti sisältää seminaarien tärkeimmät luennot. Vaikka seminaarit olivat monikielisiä, päätimme julkaista materiaalin englanniksi, koska suurin osa arviointitutkimuksesta on englanninkielistä ja keskustelua käydään englanniksi. Tällä tavoin voimme tuoda keskusteluun pohjoismaisen näkökulman.

Raportti koostuu kolmesta osasta: ensimmäisessä lähtökohtana on sosiaalialan konteksti, toisessa ne eri näkökulmat, joita tulee huomioida arvioitaessa sosiaalialan toimintaa, ja kolmannessa kuvataan eri malleja ja esimerkkejä siitä, kuinka arviointia voidaan käyttää ja on käytetty eri toiminnoissa. Miksi siis järjestää pohjoismainen seminaarisarja, jonka lähtökohtana ovat erilaiset arviointimallit? Eikö meillä ole tarpeeksi tietoa erilaisista arviointimalleista? Yleisesti ottaen kyllä, mutta käytännön työssä ei ole riittävästi hyviä malleja joiden avulla voimme yhdistää eri toimenpiteiden vaikutukset tavoitteisiin ja asiakkaiden saamaan hyötyyn. Näin voimme palauttaa tietoa toimintaamme ja kehittää palvelun laatua jatkuvasti. Seminaarisarjan perusajatus olikin antaa tietoa sellaisesta sosiaalipalveluiden ja sosiaalityön arvioinnista, jossa arviointi on kiinteä osa arkista työtä ja tapahtuu vuoropuheluna sekä käyttäjien että muiden toimijoiden kanssa. Toisin sanoen tavoitteena ei ole mallien ja menetelmien mekaaninen soveltaminen, vaan se, että

toiminnan kohderyhmä ja sisältö vaikuttavat siihen, miten arviointi tapahtuu, kuten Ernest Housen (1993) on asian esittänyt.

Arviointia tulee myös käyttää siten, että se on osa sosiaalipalvelun ja sosiaalityön kehittämistä. Kun käyttäjille annetaan mahdollisuus osallistua arviointiin, he voivat aktiivisesti myötävaikuttaa toimintojen kehittämiseen. Käyttäjien osallistumisesta sosiaalialan toimintaan on puhuttu paljon, mutta mitään käytännönläheisiä malleja ei ole olemassa. Hanne Krogstrup esittelee Bikva-malliaan (Brukar Indragelse i Kvalitetsutveckling) tässä asiakirjassa. Kyse on mallista, jota on käytetty sekä Ruotsissa että Norjassa ja nyt myös Suomessa.

Onko vaikuttavuuden arviointi utopiaa? kysyi Pekka Sulkunen luennossaan ja jatkoi pohtimalla vaikuttavuuden arvioinnin edellytyksiä yhteiskuntatieteissä. Myös sosiaalityössä on viime aikoina huomioitu näyttöön perustuvan politiikan vaatimus eli että käytännön pitäisi perustua tieteelliseen näyttöön siitä, että toiminnalla pystytään vaikuttamaan kyseessä olevien ihmisten elämään. Toistaiseksi kovinkaan paljon käytännöstämme ei perustu tieteelliseen näyttöön. Lars Oscarsson antoi luennossaan käytännönläheisiä esimerkkejä siitä, miten työn vaikuttavuutta voidaan tutkia. Hän puhui saman perusajatuksen puolesta kuin Ian Shaw (1999): että sosiaalityössä tulisi jatkuvasti tehdä pieniä tutkimuksia käytännössä ja käytännöstä.

On sanottu, että arviointi hyvin harvoin johtaa perinpohjaisiin muutoksiin, että arvioinnin käyttö välineenä on illuusio, että muutokset tapahtuvat vähitellen ja vaativat aikaa ja että arviointi on vain yksi käytännön muuttamisen alkulähde (Amba ja Stake 2001). Todelliset muutostekijät löytyvät käytännöstä itsestään. Mutta ketkä he ovat? Toisten mielestä paikalliset sosiaalityöntekijät tai poliitikot ovat tärkeimpiä muutostekijöitä, toisten mielestä taas kansalaiset ja käyttäjät. Arvioinnilla ei ehkä pystytä takaamaan parempia ratkaisuja, mutta sen avulla opimme paremmin ymmärtämään toimintaamme ja siten saamaan työkaluja, joilla voimme löytää omia ratkaisujamme ongelmiin. Toivomuksemme on, että tämä raportti antaa uusia ajatuksia, herättää keskustelemaan, tutkimaan ja arvioimaan sosiaalipalvelujen ja sosiaalityön vaikuttavuutta ja prosesseja siten, että voimme jatkuvasti kehittää sosiaalityön käytäntöjämme.

Introduktion på svenska

I april 2003 startade seminarieserien Utvärdering i Norden – nordiska modeller och metoder i Helsingfors. FinSoc planerade tillsammans med Svenska social- och kommunalhögskolan och socialpolitiska institutionen vid Helsingfors universitet samt Det finlandssvenska kompetenscentret en seminarieserie som syftade till att ge en inblick i nordiska erfarenheter av utvärdering och olika utvärderingsmodeller inom välfärdsverksamheten.

Seminarieserien bestod av tre tvådagars seminarier. Seminarierna bestod av föreläsningar av nordiska gästföreläsare, men även av finska forskare och praktiker. Det första seminariet hölls den 14-15. 4. 2003 och gav en inblick i utvärderingens framväxt och roll i välfärdsverksamheter, samt mer konkret i en utvärderingsmodell där utvärdering utgör en integrerad del av det egna arbetet. Gästföreläsare var Evert Vedung, Uppsala universitet och Göran Sandell, Göteborg universitet. Det andra seminariet hölls den 15-16.9.2003 och fokuserade på brukarmedverkan i utvärderingen och hur man med hjälp av olika utvärderingsmodeller kan få brukarnas röst hörd. Gästföreläsare var Hanne Kathrine Krogstrup från Aalborg universitet. I det sista seminariet den 10-11.11.2003 diskuterades det om hur man kan utvärdera effekter i socialt arbete. Lars Oscarsson från Örebro universitet var då gästföreläsare. Detaljerade program samt föreläsningsmaterial finns på FinSoc:s hemsida: www.stakes.fi/finsoc.

Seminariespråket var svenska och engelska, men diskussionen försiggick också på finska. Seminarierna hölls vid Svenska social- och kommunalhögskolan. Deltagande var livligt. I varje seminarie deltog ca 80 deltagare. En del av deltagarna ingick i en 10 veckors kurs i värdering som var anordnad av Palmenia fortbildningscentral vid Helsingfors universitet.

Det här working papret innehåller de centrala föreläsningarna som presenterades på seminarierna. Även om seminarierna var flerspråkiga valde vi att ge ut materialet på engelska. Största delen av utvärderingsforskningen är engelskspråkig och diskussionen förs på engelska. Det här är ett sätt att kunna bidra i diskussionen med ett nordiskt perspektiv.

Rapporten är indelad i tre delar: den första delen tar utgångspunkt i den kontext som välfärdssektorn utgör, den andra i de olika perspektiv som bör uppmärksammas när man utvärderar välfärdsverksamheter och en annan som mer konkret beskriver olika modeller och exempel på hur utvärdering utförts och kan utföras inom olika verksamheter. Varför då ordna en nordisk seminarieserie och en serie som tar sin utgångspunkt i olika modeller? Har vi inte tillräcklig kunskap om olika utvärderingsmodeller? Generellt sett, men ändå saknas det modeller inom de ordinarie välfärdsverksamheterna, som kan koppla genomförda insatsers effekter till måluppfyllelse, brukarnytta och välfärdens syfte och som kan återföra utvärderingens kunskaper till verksamheten för kontinuerlig kvalitetsutveckling, såsom Göran Sandell har uttryckt det. Det är denna grundtanke seminarieserien tagit fasta på, att få insikt i utvärdering av social service och socialt arbete där utvärdering utgör en integrerad del av arbetet och som görs i dialog med både brukare och andra delaktiga. Det är således ingen mekanisk tillämpning av modeller och metoder vi eftersträvar, utan såsom Ernest House (1993) har påpekat, att målgruppen och innehållet i verksamheten har inverkan på hur utvärderingen utförs.

Utvärderingen bör även utnyttjas så att den ingår som en del av det allmänna utvecklingsarbetet. Genom att integrera brukare i utvärderingen kan brukare själva aktivt medverka i utvecklingen av verksamheterna. Det finns mycket retorik kring brukarmedverkan i välfärdsverksamheter, men konkreta modeller saknas. Hanne Krogstrup presenterar sin Bikva-modell (Brukar Indragelse i Kvalitetsutveckling) i detta paper, en modell som fått spridning i både Sverige och Norge och nu också i Finland.

Att utvärdera effekter, är det en utopi? frågade sig Pekka Sulkunen i sin föreläsning och fortsatte att begrunda de villkor som krävs för att kunna utvärdera effekter inom samhällsvetenskaper. Frågan om evidence-based policy, att praktiken bör grunda sig på vetenskapliga bevis på att verksamheten har effekter i berörda människors liv, har uppmärksammats den senaste tid också inom socialt arbete. Än så länge baserar sig inte mycket av vår praktik på vetenskapliga bevis. Lars Oscarsson gav i sin föreläsning konkreta exempel på hur man inom praktiken kan studera effekter och pläderade för en grundidé som även Ian Shaw (1999) förespråkat, att man inom socialt arbete gör små undersökningar om och i sitt arbete.

Det har sagts att utvärdering mycket sällan är kapabel att skapa radikala förändringar, att instrumentell användning är en illusion, att förändringar sker gradvis och över tid och att evaluering är bara en av källorna till förändringen av praxis (Amba och Stake 2001). De verkliga förändringsagenterna finns inom den praktiska verksamheten. En del anser att lokala socialarbetare eller politiker är de viktigaste förändringsagenter medan andra anser att medborgarna eller brukarna är de viktigaste. Genom utvärdering kan man kanske inte åstadkomma lösningar, men den hjälper att bättre förstå vår insats och härigenom få redskap till att skapa egna lösningar på problemen. Det är vår förhoppning att vi med detta working paper kan inspirera till att diskutera, studera och utvärdera både effekter och processer i våra verksamheter för att kontinuerligt utveckla välfärdspraktikerna.

INDEX

INDEX of Figures

1 SETTING THE CONTEXT

Introduction

Ilse Julkunen

In April 2003 a seminar series on evaluation in the Nordic countries with a particular interest on models and methods started in Helsinki. The series was planned by FinSoc together with the Swedish School of Social Science and the social policy unit at Helsinki University and det Finlandsvenska kompetenscentret inom det sociala området (the Finnish-Swedish regional competence center). The main aim was to gain insight into Nordic experiences of evaluation and different evaluation models within the welfare sector.

The seminar series consisted of three 2 day seminars with guest lectures from the Nordic countries, and also by Finnish researchers and practitioners. The first seminar was held on the 14–15th of April 2003, which focused on the development and role of evaluation in the welfare sector. A particular emphasis was placed on introducing a concrete evaluation model where evaluation is an integrated part of the practical work. Guest lecturers were Evert Vedung, Uppsala University and Göran Sandell, of Göteborg University. The second seminar was held on the 15–16th of September 2003 with Hanne Kathrine Krogstrup of Aalborg University as guest lecturer. This seminar focused on user evaluation and how through different evaluation models, users are enabled to participate in practical ways in evaluating the service. In the last seminar on the 10–11th of November the general theme was how to evaluate effects in social work and social service. Lars Oscarsson from Örebro University was the guest lecturer. Detailed programs including handouts are displayed on the FinSoc web page:
www.stakes.fi/finsoc.

The seminar language was Swedish and English, but discussions were also held in Finnish. The seminars took place at the Swedish School of Social Science at Helsinki University. On average 80 participants attended the seminars and engaged actively. A group of the participants were taking part in a 10-week study unit on evaluation arranged by Palmenia at Helsinki University.

This working paper consists of the central lectures that were presented at the seminar. Although the seminars were multilingual we chose to publish this material in English. A major part of the evaluation research is published in English and most of the academic discussions are in English. This is a way to contribute to the discussion within a Nordic approach.

The report is divided into three parts. The first takes its point of departure from the context of the welfare sector; what are the general evaluation models in the welfare sector and how have they developed, and what are the risks and possibilities for projects in developing the welfare sector? The second part comprises the different perspectives to take into consideration when evaluating the welfare sector; what does evaluation mean in the context of the welfare sector with special regard to the social work being done? What are the prerequisites that we need to consider when we are evaluating

the process and the effects? In the third and final part, different evaluation models are presented which have been and can be successful when looking for knowledge and evidence in social work.

Why, then, arrange a Nordic seminar with special interest in models and methods? We have already a broad spectrum of different evaluation models, but still we lack knowledge of models that offer the potential to examine the outcomes and benefits of interventions in people's lives, which are able to assess the actual benefits and reflect them to the objectives of the intervention, and thus gain information for use in qualitative improvement to the benefit of the people concerned, as Göran Sandell has put it. It is on this basic idea that the seminar has been founded, to gain insight into the evaluation of social services and social work where evaluation is an integral part of the welfare work and which is performed in dialogues with the users and other stakeholders. Hence, it is not a mechanical assessment of models and methods we are retrieving, on the contrary, as Ernest House (1993) has also emphasised, the models and methods we choose to use should reflect the target group and the context surrounding them.

Evaluation should also be used so that it is included in the general development of the welfare sector. By integrating users in evaluating the welfare services, users may actively involve themselves in developing the services. There is a lot of rhetoric in user involvement. The current interest in more user-centred services can be traced to the transition of the welfare state and the urge to reshape the welfare services. The state supervision has diminished and social services are locally based and created on a welfare-mix principle. This creates demands for new forms of guarantees. User involvement is, thus, considered important in ensuring quality of services, in developing existing services and shaping new forms of services. Still, at the practical level, the user perspective is exploring its forms, and municipalities are uncertain of the "how and what" of user involvement. Hanne Krogstrup presents her Bikva-model (Bruger Indragelse i KVAlitetsutveckling) in this paper, a model that has been disseminated and tried both in Sweden and Norway and now also in Finland.

To evaluate effects, is that a utopia? asks Pekka Sulkunen in his lecture, where he continued by dwelling on the circumstances that we need to consider when effects are being evaluated in social sciences. The issue of evidence-based policy—that practice should be grounded on the best available knowledge of those services that do have a positive effect on people's lives—has also been debated lately within social work. Much of our practice is not grounded in scientific knowledge on effects, and very often the main obstacle is knowledge about the ways to evaluate effects. Perhaps we are too ambitious and want to grasp the whole picture? Lars Oscarsson gave, in his lecture, concrete examples of how to study effects within practice, and pleaded for a basic idea—also conveyed by Ian Shaw (1999)—that small studies should be made on and within social work.

It has been claimed that evaluation is seldom capable of creating radical changes, that an instrumental use of evaluation is an illusion, that changes happen stepwise and over time and that evaluation is only one of the sources to change practice (Amba and Stakes 2001). The real change actors are the ones who are working within the practice. On the other hand, who are they? Some say that social workers are the most important actors, others politi-

cian, whereas some claim that citizens and users are the most important actors.

Evaluation may not perhaps guarantee better solutions, but it helps us understand our activities better and by this obtain tools to create our own solutions to the problem. It is our hope that this working paper can inspire discussions, studies and evaluations of both outcomes and processes in our welfare services with the objective to continuously develop our welfare practices.

Evaluation Models and the Welfare Sector

Evert Vedung

> *Think on it constantly, think on it constantly, until the answer comes.*
> Herbert A Simon, Nobel Prize Winner in Economics

In classic public sector evaluation, three issues were considered crucial. Have the goals of public intervention been achieved out there in the real world (value assessment issue)? Is what actually transpires in the target area in any way a product of the intervention (effects issue)? And thirdly, are the answers to the first two questions useful for or used by the core decision-makers in charge of the intervention (utilisation issue)?

When evaluation began to evolve in Sweden in education policy in the 1950s, later to appear as an innovation in the United States around 1965, the three issues were addressed in very distinctive ways. The favoured solution was to commission university-based researchers to carry out the evaluation to ensure the use of the best possible scientific methodology. The academics should tease out the intervention goals as competently and objectively as possible, then measure the fit between goals and results, and appraise discrepancies if any (value assessment issue). The effects issue was tackled in a similar characteristic fashion. The optimal design, the Cadillac of program evaluation, was the randomized experiment. Professors and their associates should randomly create two equivalent groups, one an experimental group and one a control group. Then the researchers would administer the intervention to the experimental group but not to the control group. The actual development in the target area for both groups would be measured scrupulously before and after the intervention. Should there occur any differences between the groups in the post-intervention measurements, these differences could be attributed to the intervention since all other factors but the intervention were equivalent due to randomization. All this done, the findings of the evaluation were communicated by the researchers to relevant decision-makers, who were supposed to utilise the findings instrumentally, i.e. in actual decision-making and action (utilisation issue).

These views are still around in evaluation discourse and practice. But they have been pushed drastically into the background. Everyone admits that public interventions can and should be evaluated against many other value criteria than preordained intervention goals. One might evaluate against stakeholder considerations, or client wishes and expectations, just to mention two. In the evaluation of research and higher education, a special assessment tradition has evolved according to which the scientists and teachers to be evaluated first carry out self-evaluations, then academic peers are called upon to make independent assessments based upon the self-evaluations, but also upon what they have learnt through autonomous readings and dialogical encounters with those evaluated during site visits. The merit criteria used are those of the pertinent profession, which are to some extent unwritten and tacit.

In addition, a transition has taken place from evaluation as academic research, where evaluators observe the evaluatees without really interfering with them, towards evaluation as a participative, deliberative and dialogue-based exercise between evaluators and evaluatees. Evaluation should also be performed by ordinary clients, citizens, and lay people. Not only the academic person but also the common person should carry out evaluations. The justification for this direct-democratic trend, particularly in the evaluation of public service provision, such as social work and social welfare, is grounded partly in a critique of representative democracy and partly in a concern for evaluation as learning. In public service provision there is also a trend back to evaluation as academic research that goes under the battle cry of evidence-based social work or evidence-based social medicine.

Purpose

The short overview to be provided in the present chapter covers a broad landscape and some important evaluation models drawn from the general field of evaluation. Yet, the survey is admittedly sketchy; each model is only briefly touched upon and complexities are omitted. In addition, the point of departure is consciously narrow, with models classified only on the basis of the fundamental value criteria to be used to assess the merit of the activity.[1] Of course, there are many other relevant and fruitful ways of ordering general approaches to evaluation, for instance, formative-summative, monitoring-impact assessment, internal-external, ongoing-discrete, participant-nonparticipant, and experimental-statistical case study designs to attack the outcome-effects problem. And most importantly in this context, the applicability of the general models to the particular fields of social work and social welfare is mostly left to the imagination and ingenuity of the reader. Yet, it is my hope that the exposition will stimulate social work and social welfare readership to fresh insights and parallel thinking.

Taxonomy of general evaluation models

A taxonomy of evaluation models, organized according to the fundamental value criteria to be used, is presented in Figure 1.1.

[1] A Swedish draft of this chapter (basically Vedung 2002) was presented at the Nordic Evaluation Seminar, convened by Dr Ilse Julkunen, at the Swedish School of Social Science, Helsinki University on April 14-15 2003. Comments brought forward on a later English version at meetings with Riitta Haverinen, Ilse Julkunen, and Tuija Lindquist on February 9 and with Ilse Julkunen on February 24 2004 at the Finnish Evaluation Unit for Social Services (FinSoc) in the National Research and Development Centre for Welfare and Health (Stakes) have been incorporated. The present report represents a drastically abbreviated and reworked but also to some extent extended rendering of chapter 4 in my book *Public Policy and Program Evaluation* (New Brunswick, N.J.: Transaction Publishers, paperback 2000, www.transactionpub.com/cgi_bin/transaction-publishers.storefront). Revisions have been undertaken in all sections. Procedural models, particlarly deliberative democracy, are added. Economy has been adjoined to productivity and efficiency. Footnotes have been kept to a minimum due to time pressure.

I Substance-only Models
- 1 Goal-attainment model (Effectiveness Evaluation)
- 2 Side-effects model
- 3 Client-oriented models
- 4 Stakeholder models
- 5 Professional models: Peer Review, Self-evaluation

II Economic Models
- 1 Economy model (Cost-only model)
- 2 Productivity model
- 3 Efficiency models: Cost-effectiveness, Cost-benefit

III Procedural Models
- 1 Legality
- 2 Equity (Rule of Law)
- 3 Publicity
- 4 Representativeness
- 5 Participatory democracy (Public participation in final decision-making)
- 6 Deliberative democracy

Figure 1. 1 Evaluation Models

The major line of demarcation distinguishes substance-only models and economic models on the one hand and procedural models on the other. While substance-only models primarily address substantive intervention content, outputs and outcomes (and, secondarily, the processes preceding and connecting them) and economic models attend to intervention costs, the procedural models check for legality, equity, publicity, representativeness, participatory democracy, deliberative democracy, and other qualities of the processes according to which the interventions are supposed to be handled by ministries and agencies. Although of long-standing importance, procedural models of evaluation will not be dealt with here with the exception of deliberative democracy, which will be briefly covered at the end of the chapter. In the main, my survey will focus on the dominant substance-only models and economic models.

Substance-only models constitute a fairly large and varied group. In addition to the classic goal-attainment evaluation, there can also be found in this category the side-effects evaluation, client-oriented evaluation, stakeholder evaluation and professional evaluation. The peculiar property of economic models is that they pay heed to intervention costs. The three basic economic models are the economy, productivity and efficiency models. The economy model is directed at the cost aspect of public interventions and nothing else. Productivity and efficiency models, on the other hand, integrate cost and substance aspects of public interventions.

My condensed exposition will start with substance-only models, proceed with economic models, and end with one procedural model, deliberative democracy.

Goal-attainment Evaluation (Effectiveness Evaluation)

In social welfare evaluation as in evaluation in general, goal-attainment evaluation is an approach of long standing. The two basic ingredients of the goal-attainment model are goal-achievement measurement and intervention impact assessment. The key question in goal-achievement measurement is: Are the results in accordance with intervention goals? And the impact assessment issue can be formulated: Are the results at least to some extent produced by the intervention?

The plain anatomy of goal-attainment evaluation is outlined in figure 1.2.

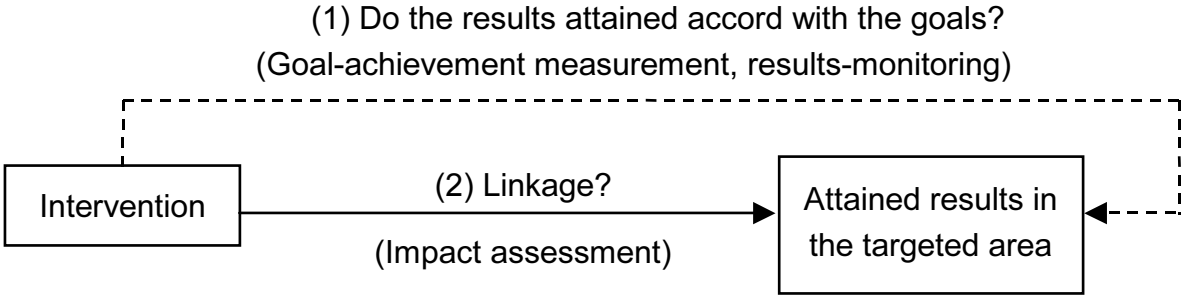

Figure 1. 2 Goal-attainment Evaluation (Effectiveness Evaluation)

Goal-attainment evaluation is a substance-only model because it restricts its questions primarily to the substantive content, output and outcomes of the intervention (and, secondarily, processes preceding and connecting them). It differs from economic models in that it pays no heed at all to intervention costs.

The Strength of the Goal-attainment Model

In earlier literature, the goal-attainment model reigned supreme. Public sector evaluation was goal-attainment appraisal, period. Since the 1970s, however, goal-attainment evaluation has been under constant attack. Practically every author who wants to advance some new idea in evaluation starts with a strong polemic against the goal-attainment model.

There are, however, several worthwhile reasons in favour of goal-attainment assessment in the public sector and in social services. In this context I shall mention only one: the argument from representative democracy.[2]

In a democracy, all power belongs to the people. Yet, the people cannot make all the complicated decisions concerning the well-being of citizens due to a lack of competence and lack of time. The people do not have time to participate in the hundreds of thousands of decisions. And the people do not have the necessary competence to make wise decisions on, for instance, placement of patients in line for surgery, or day-to-day care for ailing senior citizens in public-sector homes for the elderly. For these reasons the citi-

[2] Several other reasons are presented in Vedung 2000 (1997):40 ff.

zenry must elect political representatives to make the decisions for them. But representatives in political assemblies don't have time and competence to make all decisions. They must delegate their power to governments to make decisions for them. But governments don't have time and the specific knowledge necessary, so they in turn have to delegate to civil servants and professionals to take decisions, etc. etc. The public sector is made up of long chains of principal-agent relationships.

To safeguard that principals are not cheated by their agents, public-sector goals set by political assemblies are circumscribed by formal rule systems to an extent that has no counterpart in, for example, corporations, voluntary associations, or families. Political leaders are supposed to honour the rules of the constitution, and the rules of procedure in popular assemblies. Once a decision on goals, policy instruments, and services comes out of this system, it has a status that cannot be compared to decisions in other social bodies.

This is also the case with other entities in the public sector such as governments and agencies, acting on delegation from assemblies entrusted with decision-making authority. If an agency adopts a program in order to reach some goals, these goals get their legitimacy from the fact that the agency's decision-making authority has been delegated to it by the government, and that the government, in turn, has received its authority to do so from parliament, and the parliament, in turn, from the people. It is a merit of the goal-attainment model that it recognizes this democratic aspect of public-sector goals, which is often overlooked by students of general organizations.

In sum, the goal-attainment model scores an important point with respect to its tilt toward representative democracy and the parliamentary chain of control. On the other hand, the goal-attainment model also suffers from persistent flaws and weaknesses.

The Shortcomings of the Goal-attainment Model

The most significant general reasons against the goal-attainment model are[3] the haziness argument, and the unintended side-effects argument. The haziness argument maintains that intervention goals are deficient as criteria of merit due to their obscurity.

There are two kinds of goal obscurity: goal indeterminateness and goal catalogues. Occasionally, programs are based on indeterminate goals. Particular goals may be ambiguous and carry two or more simultaneous meanings. Yet ambiguity in this sense of dual meanings is exceptional in political and bureaucratic language, and barely bothers evaluators. More uncertainty is caused by vagueness. A goal is vague if it does not delineate clearly cases where it is or is not applicable. The outer border delimiting the extension of a vague word is so fuzzy that within a certain range it is impossible to know what is included in the extension and what is not. Rampant in political rhetoric, vagueness is one favourite expedient to settle political conflicts through semantic formulas without really resolving them.

[3] For a more extended exposition of the arguments, see Vedung 2000 (1997): 43 ff.

The second major obscurity is produced by goal catalogues. Most large social reforms contain impressive directories of diverse goals. While a single goal may be hailed as the major one, it is also often maintained that this one must be balanced against all the others, maybe including potentially conflicting ones. But the necessary trade-offs between the various goals are not indicated, which makes it impossible to elicit from such lists of goals one distinct, transparent, expected outcome. Thus, program goals do not offer any safe guidance for continued data assembly. They are not lucid enough to be usable as value criteria against which to measure intervention successes, shortcomings and failures.

The goal-haziness argument reveals an important misfit between the requirements of the goal-attainment model and the way public policies, programs, and activities are often composed. If elected officials and program planners have not specified individual goals into measurable objectives, and if they have not balanced the various stated goals into one global outcome or output measure, the goal-attainment evaluator cannot summarize her findings into a completely value-neutral evaluative judgment. She can do so only after she has clarified the goals and prioritized among them in a fashion that will cast doubts on the objectivity of the whole enterprise.

The third general counter-argument, about unintended side-effects, is in my view the crucial one. Public sector interventions, in social security and social work for instance, invariably lead to consequences which were not foreseen in the original decision situation. Were the evaluators to confine themselves to ascertaining the achievement of premeditated program goals, the search for serendipitous results or unanticipated side-effects outside the goal area would not be included in the evaluation process. The evaluation findings would exhibit a tunnel vision of events, and produce a biased, if not fundamentally wrong, picture of what the program has attained. In all likelihood, a program generating some interesting spin-off effects must be better than a program producing several undesirable spillovers.

In sum, the major strength of the goal-attainment model is grounded in the theory of representative democracy. From a representative government perspective, such principals as citizens and elected politicians need goal evaluation to check whether their agents actually carry out what they are obliged to do.

However, the goal-attainment model is also open to several forceful objections. It has some problems with hazy goals, and goal catalogues, pervasive as we all know, in public policy. The most compelling rebuttal, however, emanates from the model's blindness to side-effects.

At this point, I would like to present a model that expressly considers the weighty side-effects argument, while retaining the fundamental goal-orientation of the goal-attainment model: side-effects evaluation.

Side-effects Evaluation

The side-effects approach is similar to goal-attainment evaluation to the extent that intervention goals are retained as the fundamental value criteria. The novelty with side-effects evaluation is that the hunt for targeted results is supplemented by a search for by-products outside the target areas.

The fact that the side-effects model, like the goal-attainment model, is based on goals emerges from the expression 'side-effect'. Side-effects must be defined in relation to the intended main effects. Main effects' can be defined as those actual, expected and wanted consequences, which at least partly are produced by the intervention. Consequently, main effects are associated with the substantive objectives of the policy-makers and with what they believe the intervention is capable of achieving. Furthermore, main effects are by definition anticipated as well as positively valued by the policy instigators. From the interventionist point of view, the term 'side effect' can be defined as at least a partial consequence of the intervention occurring outside the intervention target areas.

The idea underlying the side-effects model is that public interventions may produce things other than intended results. They may lead to great positive surprises but also generate new problems.

The basic skeleton of the side-effects model is exhibited in figure 1.3.

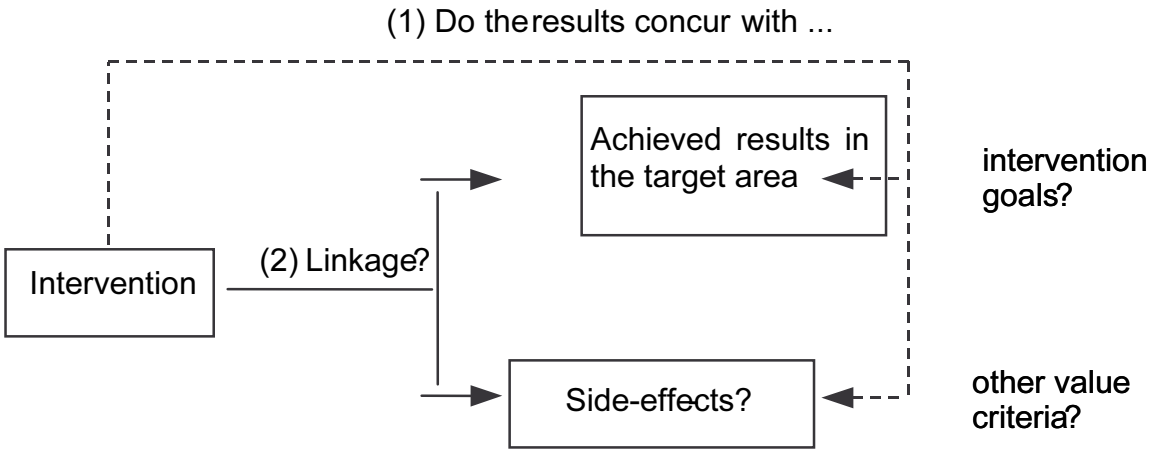

Figure 1. 3 Side-effects Evaluation

Public interventions may also create perverse effects. Perverse effects run exactly counter to the ones intended or expected, for instance by the intervention instigators. These impacts may occur in the target area or areas of the public intervention. They may also occur outside the target areas and be side-effects. Perverse effects may occur far down in the purported chain of control, i.e., in the second, third, or even fourth outcome stage. In addition, they often crop up after many years.

Perverse effects are also different from null effects. Null effects means that interventions produce no impacts at all inside or outside their targeted areas.

If perverse effects and null effects occur in the targeted areas, the goal-attainment model with all its attention directed at what happens in these particular fields has no problem with handling them. But this also means that the model cannot discover and ascertain side-effects because they crop up outside the targeted areas.

Side-effects can be anticipated and considered in calculations preceding decisions to adopt policies. They may be beneficial as well as detrimental. However, some side-effects are no doubt unanticipated as well. Like their

foreseen counterparts, they might be felicitous or deleterious. Increased radon radiation in dwellings, a probable consequence of better insulation causing less draft, may be adduced as an example of a deleterious, unanticipated side-effect when the Swedish housing insulation scheme was adopted in the 1970s.

My argument on main effects, side-effects, perverse effects, and null effects is summarized in the effects tree in figure 1.4, showing which aspects of effects that might be studied in evaluation research.

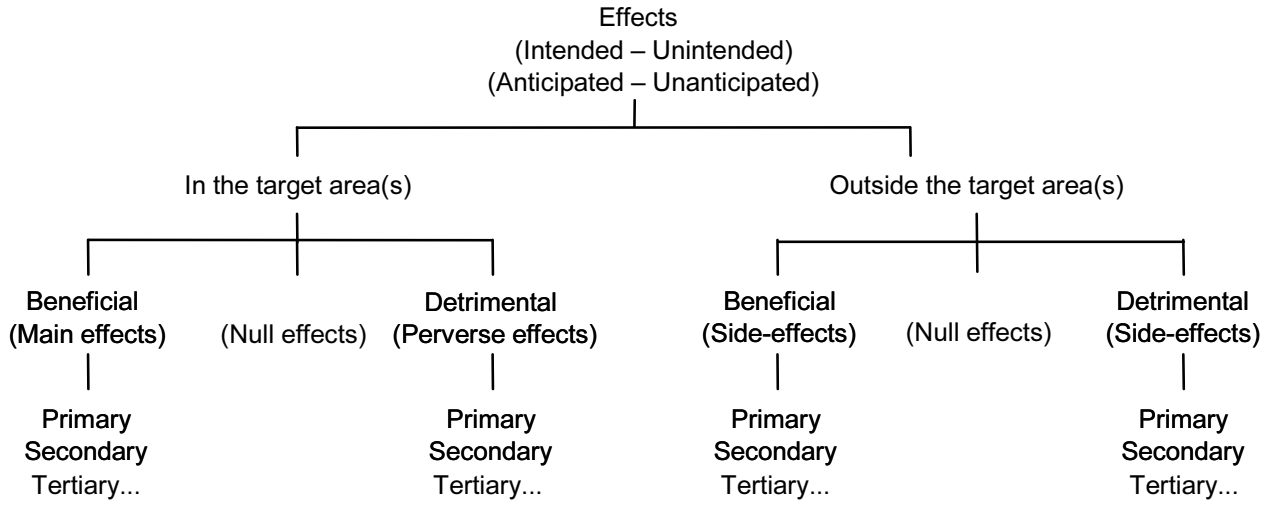

Figure 1. 4 Main Effects, Side-Effects, Perverse Effects and Null Effects

The importance of noticing perverse effects should be obvious to every evaluator. If the intervention, a piece of social work for instance, produces consequences contrary to its main stated purpose, there must be something wrong with it. But why is it so vital to pay attention to side-effects? Because by-products, whether detrimental or beneficial, are crucial factors in every inclusive judgment of the operation of an intervention. Should it turn out that side-effects, which have been known, discussed, and positively valued in advance, have not materialized in spite of the fact that the intervention has been on the books for the intended period of time, this ought to be considered in any appraisal of the intervention.

I strongly prefer side-effects evaluation to goal-attainment evaluation. Indeed, the major rationale for doing public sector evaluation in the first place is that state actions to some extent are unpredictable and regularly result in either side-effects not originally foreseen or no side-effects originally foreseen and wanted. It is an important duty of evaluation to map and assess what is happening with side-effects.

Client-oriented Evaluation

Client-oriented evaluation employs the desires, values, objectives, concerns, expectations, and assumptions of the intervention addressees (intervention

participants, intervention recipients, intervention targets, service-users) as its criteria of merit. At the heart of the full-blown client-oriented model is the concern whether the intervention, its implementation, outputs, and outcomes satisfies client desires, values, and interests, or are in accord with client expectations, and assumptions. The experiences of the intervention users are absolutely central to the client-oriented models.[4]

Several approaches might be considered client-oriented. They may be ordered according to the degree of client involvement. There are considerable differences between them.

Think about an example concerning client satisfaction with public-sector services. Assume that in the city of Helsinki in the year 2000, 50% of the parents with children in municipal daycare centres were satisfied with the daycare of their children. This constituted a decrease from 1999 when 55% were satisfied. The plunge obviously worried the city councillors; in a city council debate in January 2001 the members expressed their concerns about the situation. The city council unanimously maintained that parent satisfaction must increase; therefore, the councillors decided to set 70% parent satisfaction as the goal to be achieved in 2003. In the beginning of 2004 a new study demonstrated that in 2003 60% of the parents were satisfied. Certainly both the 2000 and the 2004 study is client-oriented. Are they client-oriented evaluations? The answer is yes. User satisfaction is a client-oriented value criterion. Interestingly, the second study might be considered a goal-attainment evaluation as well, since 70% parent satisfaction was the goal of the city council.

There are more advanced forms of client-oriented evaluation, where the clients are much more involved. We may think of circumstances under which the evaluation is initiated, funded, designed, and carried out by the clients themselves. This is very different from a situation where the evaluation is initiated, funded, designed and carried out by some higher authority, for instance by politicians, administrators or service delivery personnel. Let me reason from the case where the evaluation (i) is commissioned by the administrators, and (ii) is planned to involve the service-users much more than merely asking them in a questionnaire about their service satisfaction.

The primordial step in the practical application of the more elaborate client-oriented evaluation model is to create an evaluation team and determine its terms of reference. Next, the team locates the intervention clients. Since the evaluation normally cannot cover them all, a sample from the target population must be picked. Now, the evaluators may tell clients that they may pass judgments on some aspects of the service. Ideally, the clients themselves are encouraged to select the aspects on which to pass judgments. For instance, clients may judge program output, service availability, service quality, or even service process and service administration. Is the core service tailored to meet the most detailed demands of the clients? Are the en-

[4] Actually, there is also a school of thought which uses client *needs* as point of departure for client–oriented evaluation. Some argue that needs can be traced and ascertained scientifically. Other maintain that needs can be determined by politicians or professionals; these needs may differ radically from the clients´ own desires and expectations. To simplify I have avoided the needs issue here.

counters of the clients with the service employees respectful? These are two questions that might be answered by the service-users in their self-reports.

The clients may also choose to raise the causal issue, that is, estimate service impacts on themselves or on the client community in general. In such client-outcome evaluation, targets try to determine the relative change in themselves or in the overall client body as a result of their participation in specific treatment modalities. In technical terms, clients compare what actually has happened with the intervention/treatment in place to what would have happened had there been no treatment or a different treatment. Client-outcome evaluation uses the so-called shadow controls design to make an assessment of intervention impact.

In data assembly, advocates of client-oriented models prefer client self-observation and sustained interviewing to questionnaires, documentary methods, and evaluator observation. In-depth interviewing of individual targets is one favoured technique; distribution of self-report instruments that clients and their network can easily complete themselves is another. In some cases, client-oriented evaluators endorse focus-group interviewing which allows for group deliberations among the participants and between the participants and the evaluator. The evaluator tries to create forums of debate to promote deliberative richness. This will support the development of new ideas, service concepts, solutions and technologies. It might also, as a side-effect, educate participants to become better citizens in the future.[5]

It has taken an inordinately long time before the clients were recognized in public administration. The notion that the amount, kind, and quality of social services are, and should be, determined by elected officials or by other organs acting on delegation from elected officials has exerted a commanding influence over administrative thought. Representative democracy and professionalism have been the predominant modes of thought in the public sector.

Today, client-oriented models are employed in numerous evaluative contexts, particularly those concerning public service provision such as child care, nursing homes for the elderly, public housing, mental health, urban transit, public utilities, parks and recreation, and physical health services, where clientele participation is crucial to the operation of the program. Client-oriented models are used to evaluate library services, arts, zoos, and museums. It is a favourite with educators. At universities, students are routinely requested to share their opinions of courses, reading lists and lectures. They are asked to rate their teachers´ abilities to organize the course contents, to stimulate and promote altercation and discussion, to stir student interest, motivation and critical thinking, and to show concern and enthusiasm for the students. At American universities, these evaluations are occasionally used to rank faculty and courses from a student perspective so that future prospective students can make better-informed choices.

[5] I shall not dwell here upon how the findings from a round of client-model evaluation can be brought to another stakeholder group, the reactions of which in turn will be brought to still another stakeholder audience, and so on. The so–called BIKVA-model will be dealt with in Hanne Krogstrup´s chapter in this volume.

Pros and Cons of Client-oriented Evaluation

Client-oriented models are justified in several ways. First, some philosophers ground them in political ideologies based on the notion that public administration produces goods and services for customers in the market place, and maybe also on the superiority of the market place over public-sector provision. They claim that customer pressures expressed through attitudes and suggestions for improvement towards evaluators and service providers will lead to the improvement not only of the core service, but also of service processes, service delivery, service effects, and customer satisfaction. In buying a commodity in the store, the consumer pays no attention to producer goals and purposes. Her own assessment of the value of the good is what counts. This should also be the case in public sector social services and social welfare, because then the services would become more clearly geared toward the wishes and expectations of the users.

On the other hand, for the second type of client-oriented models, a direct-democracy and deliberative-democracy case can also be made. According to this justification, the customer parallel cannot be pushed too far, since the client notion includes a participatory as well as a deliberative aspect, which is absent from the customer concept. The participatory feature suggests that clients are also citizens who may voice their complaints and desires to the evaluators and service providers, and to some extent influence and take responsibility for service content. The deliberative feature engenders a discursive, reasoning, discussing, learning-through-dialogue countenance, which may educate clients to become better citizens in general: the consumer as citizen rather than the consumer as customer.[6]

Client-oriented evaluation may also increase the legitimacy of the intervention. If clients are asked for their opinions, allowed to participate, and have some influence on the intervention, their acceptance of or support for the intervention will probably increase. In similar fashion, evaluation may also improve the image of the service providers, that is, increase imaginary values in addition to the value of the core product.

In addition, client-oriented evaluation may foster effectiveness and efficiency because concentration on clients may force service-providers and managers to do away with many other things they can be preoccupied with besides providing good services.

But evaluators must be aware of the tendency of the clientele to exaggerate complaints in order to get more service. Clients may also nurture fiscal illusions. Greater client involvement in evaluation may surrender power to groups with vested and narrow interests.

In my view, the client-propelled models may supplement the previously presented approaches, since they pose other problems for consideration. They can make important contributions to evaluation, but should not be allowed to replace the other models. The requirement that the civil service must be responsive to client concerns is sound, but only within limits. It can

[6] My extensive experience as an evaluation instructor has taught me to emphasize the difference between intervention clients (intervention participants, addressees) and evaluation clients (evaluation users). The evaluation client is the person, group or agency that has commissioned the evaluation of an intervention or is supposed to use its findings, whereas the intervention client is the intended or actual recipient of the intervention.

never take precedence over the requisite that front-line operators should follow the directives of their hierarchical administrative superiors, and indirectly political bodies like Parliament, the Municipal Council, and, ultimately, the citizens, whose votes have determined the composition and general policy direction of these bodies. The elderly in a community who enjoy municipal social home aid cannot take decisions that run counter to the rules of the agents and principals in the representative chain of control. They cannot unilaterally lower the service fees, for instance. Evaluation models grounded in representative democracy must take precedence over client-oriented models. Client criteria are reasonable to use, but within limits; they must be balanced against other criteria like goal-attainment and professional norms for service excellence.

Stakeholder Models

The merit criteria of stakeholder evaluation are the concerns and issues of those actors who have an interest in or are affected by the intervention. This is quite different from using prefixed intervention objectives as criteria of worth—as in goal-attainment evaluation or even side-effects evaluation. Stakeholder evaluation, however, does resemble the client- and profession-oriented models, the major difference being one of scope: while client- and profession-driven models are basically concerned with one category of affected interests, the stakeholder model is geared to all of them. Figure 1.5 displays a survey of potential stakeholders in local social welfare interventions.

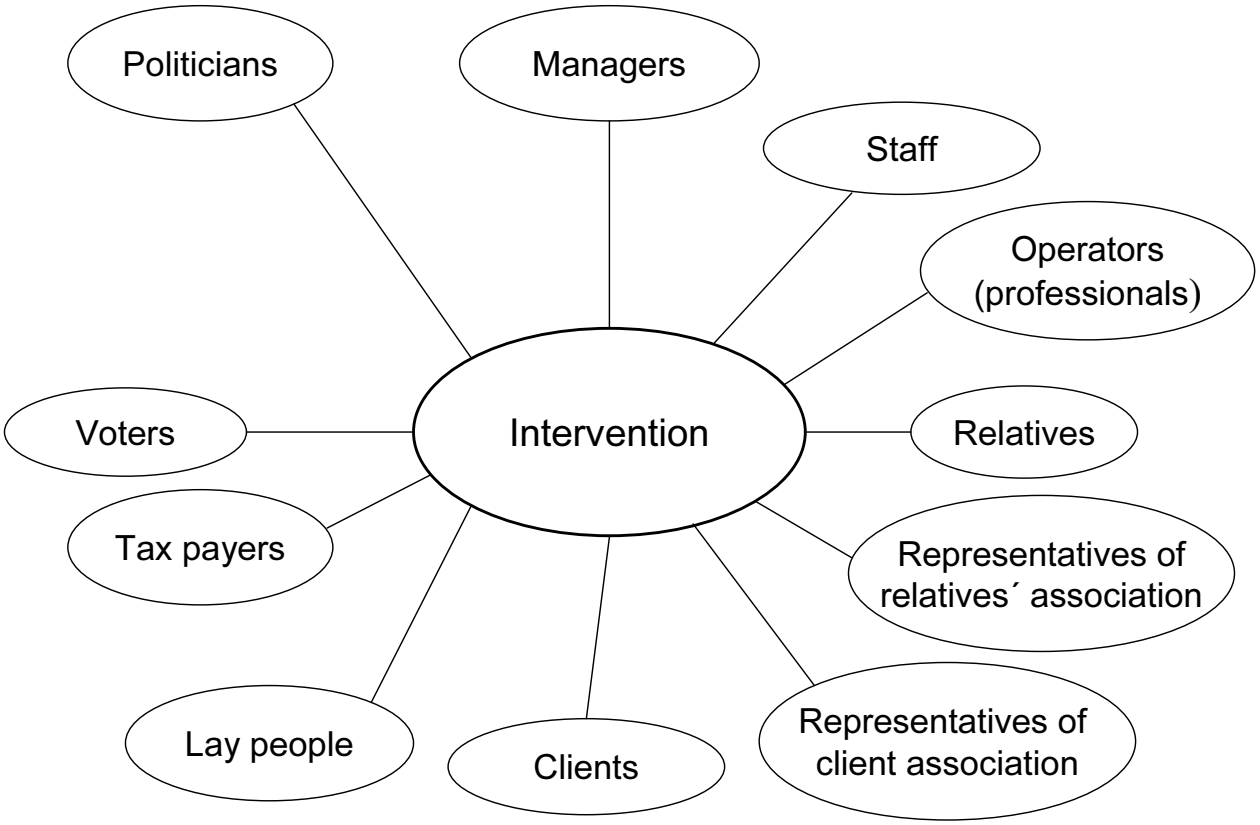

Figure 1. 5 Potential Stakeholders in Local Social Welfare Interventions

Stakeholder evaluation can proceed in different ways. The stakeholders may constitute themselves as the evaluation team and carry out the evaluation. The evaluation may also be conducted by particular evaluators, who elicit the views of the stakeholders. In the sequel, I shall reason from the case that stakeholder evaluation is carried out by particular evaluators.

Stakeholder-based evaluations start with the evaluator mapping the major groups who are involved or thought to have an interest in the emergence, adoption, execution, and results of the intervention. The evaluator identifies the people who initiated, hammered out, funded, and adopted the intervention, i.e. basically the politicians. She identifies those who are charged with its implementation: senior, middle, and junior managers, staff, and front-line operators, who actually deliver intervention output. She singles out the intervention's primary target group, the clients, and the clients´ associations. She identifies relatives and relatives´ associations. She may also include lay people. And she searches for those who know they have a stake in the intervention but prefer to keep a low profile and those who are unaware of the stake they hold.

Advocates of stakeholder models nurture a strong predilection for qualitative, interactive methodology. One key expression is interactive search procedure. The evaluator must talk to the stakeholders to elicit their narrative histories and observational data, which in turn should be allowed to affect the evaluator's next step in the search procedure. After a while, she

might discover both the purported and the genuine aims of the intervention, and what concerns various stakeholders nurture regarding it. With time, the evaluator gets more involved and can start to determine which dimensions and concerns of the stakeholders should be included in the study. Only then can she take a stand on what the outline of the evaluation should be.

It is typical of the stakeholder model that the evaluator is permitted to search rather extensively for the dimensions and the crucial concerns on these dimensions. The idea is that the evaluator must be responsive to the issues and concerns of the affected people and let these govern the next step in the investigatory enterprise. Through interactive communication she is supposed to find out which stakeholder dimensions and concerns are to be taken seriously and probed more deeply. The evaluation design will be gradually determined. Stakeholder evaluation is responsive evaluation.

To elicit final data on stakeholder dimensions and concerns, advocates of stakeholder models prefer stakeholder self-observation and sustained interviewing to questionnaires, documentary methods, and evaluator observation. In-depth interviewing of individual targets is one favoured technique. In social service and social work, distribution of self-report instruments that clients, parents, relatives and other stakeholding networks can easily complete themselves is used. In some cases, client-oriented evaluators endorse focus-group interviewing which allows for group deliberations among the participants and between the participants and the evaluator. The evaluator tries to create forums of debate to promote deliberative richness. This will support the development of new ideas, service concepts, solutions and technologies. It might also, as a side-effect, educate participants to become better citizens in the future. After data are amassed and processed, the reporting of findings, which might vary from one stakeholder to another, will commence. The key word seems to be 'portrayals', that is, information-rich characterizations using pictures, anecdotes, thick descriptions, and quotes. The comprehensive holistic view mediated through a portrait is important. Normally, several criteria of merit, standards of performance on these criteria and several comprehensive assessments will be included.

Advantages of Stakeholder Evaluation

Stakeholder models have several advantages of which four will be mentioned here.

The democratic arguments for the stakeholder models depart from participative and deliberative points of view. True, democracy means that citizens in general elections vote for competing elites that are supposed to make decisions on their behalf (representative democracy). Yet, the citizenry should also be able to partake in final public decision-making between elections (participative democracy). Furthermore, discussion, dialogue and debate are also important democratic values because they help people to form and refine their beliefs and preferences (deliberative democracy). Stakeholder models satisfy these participative and deliberative values somewhat more than the goal-achievement model and the side-effects model.

According to the knowledge argument, it would be foolish of the evaluator to avoid the intervention insights which those involved undoubtedly

have. Stakeholders nurture convictions about inadvertent side-effects, sophisticated implementation barriers, and outright cheating which may furnish the evaluator with ideas about topics for further investigation. Since stakeholder-orientation will bring up more aspects of the subject-matter for discussion, the quality of the evaluation findings will increase. All-in-all, it is easy to agree with the recommendation that almost every evaluation ought to begin with the determination of relevant actors and rounds of interactive data assembly.

There is also a utilization argument in support of stakeholder evaluation. Findings from goal-attainment and side-effects evaluations carried out with quantitative methodology and with no interactive involvement of stakeholders seem to have little impact. The reports are buried in desk drawers and heaped on the bookshelves of the public authorities involved, unread and forgotten. This seemingly irrational behaviour has puzzled and exercised researchers. Why do decision-makers behave in such an unwise fashion? The allegedly major explanation is that evaluators work in splendid isolation, with too little communication and interaction with prospective users. The stakeholder approach increases the chances that issues of genuine interest to concerned parties will be addressed. It brings to light information that meets the real requirements of the different stakeholders, thereby enhancing the probability that the findings actually will be put to use.

A last rationale for stakeholder models is that they might promote compromises, and forestall bitter political struggle. Stakeholder assemblies are consensus-building mechanisms. They are vehicles for shaping agreement on the results of earlier efforts, and most importantly, proposals for future action. Consensus-building and the rendition of legitimacy to fundamental decisions are considered great advantages of stakeholder evaluation.

Drawbacks with Stakeholder Evaluation

There are also obvious drawbacks with stakeholder models.

Stakeholder evaluations are inordinately impractical and resource-demanding, since every stakeholding constituency must be contacted and nurtured.

Stakeholder models are fuzzy. They provide no authoritative answer to the question of who the stakeholders are. The range of stakeholders must be decided on a case-by-case basis. Furthermore, all the stakeholding audiences, however selected, are treated as equals. But in a representative democracy, elected politicians must carry more weight than administrators or experts on the substantive matters under consideration, just to pick a few. The stakeholder model embodies no priorities among the stakeholders. There is a risk that small, well-organized and very committed stakeholders are consulted and listened to more than vaguely concerned groups and the majority of citizens.

This raises a problem with stakeholder evaluation, or, more properly, collective decision-making on the information base brought forward by stakeholder evaluations. Since representative democracy is the dominant form of democratic government in most developed societies, decisions on various sectoral policies cannot be left to the stakeholders. The power belongs to the

people, its representatives and their delegates. Stakeholder evaluation must work within the frames fixed by representative democracy.

Another objection involves the allegedly damaging risk that the stakeholder model will embrace a pragmatic theory of truth. Truth may turn into a matter of usefulness, utility or acceptability to stakeholders. Stakeholders often entertain politicized views of program effects. Supporters ascribe everything positive occurring after the program as caused by the program and everything negative as caused by something else. Detractors hold the opposite view. Facts are essentially contested. In these situations the various parties will accept only those findings that fit into their preformed opinions.

In conclusion, while hazy in its contours and quite controversial, the stakeholder model carries some important merits. Utilization and compromise arguments speak in its favour. Another strong reason for it is the knowledge argument: it can be used as a search strategy at the start of the evaluation in order to get a quick and provisional grasp of the meaning, implementation and outcome of the intervention.

Finally, a case from deliberative and participatory democracy can be made for stakeholder models. Through the use of the stakeholder approach, affected interests can deliberate public affairs and learn to become better citizens in the future. They can participate and influence the final outcome. On the other hand, this must take place within the confines set by representative democracy.

Professional Models: Peer Review and Self-evaluation

Professional evaluation models imply that members of a profession—experts as it were—are entrusted with evaluating their own or their colleagues´ performances with respect to the profession's own criteria of merit and standards of quality. Professional models are expertise-oriented approaches.

In some areas of public life, criteria of good quality are so complex and the subject-matter so intricate that political officials have found it wise to leave the shaping and debating of them to well-educated professionals. Architects, judges, professors, doctors, veterinarians and engineers would be cases in point. Hence, it is also considered natural to delegate ex nunc and ex post evaluation to the professions.

Professional evaluation is conducted in several ways. Self-evaluation is one form. Then, the professional herself evaluates her own performance or professionals in an organization together evaluate their organization’s performance. In another type, peer review (collegial evaluation), the evaluation is conducted by an external collegium, which by definition is an assembly of professional equals. Ideally, these equals should be somewhat better in their area of expertise than the colleagues they are supposed to assess. The trust of those evaluated for the evaluators is important. In this way, lawyers evaluate lawyers, scientists evaluate scientists, and surgeons evaluate surgeons. Social workers might also evaluate social workers, welfare operators evaluate welfare operators, and so on and so forth.

In the fields of research and higher education, a fairly special evaluation tradition involving professional evaluation models has evolved. It differs

very much from goal-attainment evaluation and side-effects evaluation but also from the economic models which will be covered later in this chapter. The evaluation is to some extent based upon dialogue, discussion and deliberation between the evaluators and the evaluatees and among the evaluatees themselves. It is not carried out as a scientific piece of work but as an exercise in qualitative, interactive search and appraisal.

The procedure usually starts with self-assessment by the evaluatees. The professionals to be evaluated carry out an appraisal of their own performance, of the research project, the research program, or the university department. They discuss and assess strengths, weaknesses, opportunities and threats (a so-called SWOT analysis). Then, renowned scientists of the particular field are assigned to appraise the quality and relevance. These peers base their assessment on the written self-evaluations, additional documentary evidence, and site visits with face-to-face evaluator–evaluatee dialogues. The evaluators pass their preliminary judgments in a draft report, often organized according to the SWOT scheme. Then the evaluatees are given the opportunity to comment on the draft report before it is finalized. The final peer review report is particularly aimed at performing an overall quality judgment of the evaluand.

Usually, peer review of research is—and ought to be—at least partly interactive. The evaluators meet and listen to the evaluatees and the evaluatees learn from the evaluators.

To provide an example, the evaluations of research efforts sponsored by the Swedish Council of Building Research at the end of the 1980s were assigned to take special interest in:

a. The relevance of problem selection and research design
b. The suitability of the methods of analysis
c. The tenability and validity of arguments and conclusions
d. The work in relation to the discourse in the pertinent area
e. The practical applicability of research findings
f. The worth of the research community shaped by the research enterprise
g. Agreement between original intentions and findings reached

After the directives had been written, appropriate experts were approached. Preferably, the experts should have more specialist knowledge in the field than the colleagues whose research is to be evaluated. They should also be independent; for instance, they must not have carried out research work in the area under scrutiny in cooperation with the people to be evaluated. There is an important difference between those peer reviews where the evaluatees have suggested and agreed in advance on the choice of experts, and reviews where the evaluatees have no say in peer selection.

After the expert group, the 'peer collegium' as it were, had been chosen, the upcoming work was organized. Invariably, the reviewers and the reviewees interacted with each other during the reviewing process. To the evaluators it was important to take the concerns and arguments of the evalu-

atees seriously and try to include or at least consider them in the evaluation. On many occasions, the evaluatees were asked to provide relevant publications and other research material in order for the evaluators to become sufficiently informed. The evaluators were then given time to read the material to inform themselves about the evaluatees and their products. Then, each researcher and research group was visited for presentations and informal talks. In due time, a preliminary formal report was drafted. An important feature was that the preliminary report was circulated to affected researchers whose written comments were explicitly solicited. These comments were subsequently paid attention to when the evaluators composed the final report. However, they were not published along with the finished report. In the Scientific Commission of the National Swedish Building Council, the whole procedure took approximately 18 months.

Peer reviews frequently produce questionable results. Studies with matched panels show that peers use widely different merit criteria and performance standards and reach miscellaneous conclusions. However, in technically and substantively complex fields, interactive collegial evaluation is probably the finest method available to judge the quality of what is produced.

Economic Models: Economy

Common to all substance-only models is their negligence of costs. No matter how resource guzzling a program has been, substance-only evaluators concentrate on its substantive results (and processes preceding these results) and disregard costs. Attention to costs, on the other hand, is a typical feature of the economic models of public policy and program evaluation. Economic models are commonly divided into economy, productivity, and efficiency.

The simplest model, economy, single-mindedly focuses on cost development and cost reduction. For this reason, a more telling name would be the cost-only model. Has the intervention cost less this year than the year before? Has the intervention been implemented as cheaply as possible? Is it possible to reduce costs? How can costs be lowered? Can budgetary deficits be reduced? In which ways can spiralling expenses be capped and reduced?

Economic Models: Productivity

The second merit criterion of good economic performance is productivity. Productivity is the relationship between output of products and services and input of resources. In short: output per euro spent, or output per hour worked, or output divided by input. Productivity can be expressed through the simple algorithm shown in figure 1.6.

$$\text{Productivity} = \frac{\text{Output (=value of output)}}{\text{Input (=value of input)}}$$

Figure 1. 6 Productivity

For a deeper understanding, let me provide a contrived example from the library community. In computing the productivity of municipal libraries the following two measures may be used:

$$\frac{\text{number of books borrowed}}{\text{costs in € (alt. number of hours worked)}} = \textit{Cost productivity} \text{ (alt. } \textit{work productivity}\text{)}$$

The difference between the two alternative measures is that costs in the former case are indicated in monetary terms, in the latter case in the number of hours worked, that is, as physical entities. The time unit used could be the fiscal year, the calendar year, or even a monthly period.

Other possibilities would be the ratio of library holdings (number of books kept) to costs, the number of inhabitants in the municipality to costs, or the number of borrowers to costs.

To say something worthwhile about the actual productivity of a public agency, a reference case is needed as comparison. Several reference cases are used: comparison with past performance, with similar institutions in the same country, similar institutions in other countries, goals of the political bodies, client goals or stakeholder goals.

Productivity as a measure of public sector activities has some disadvantages. Outputs, for instance, may be exacting to catch and compute, even though productivity only presupposes that they are indicated in physical, not monetary terms.

Let me return to the municipal libraries to illustrate the difficulties of finding valid output indicators. Is the number of borrowed books really a relevant and exhaustive output measure? Admittedly, it is relevant. To provide the public with an opportunity to borrow books is reasonably the most important task of a public library. But it is certainly not exhaustive. A Finnish report concludes that only 30-50 percent of the library clients borrow books. The other patrons visit the library to read newspapers, magazines and journals. They frequent the reference library to use dictionaries and encyclopaedias or the music department to listen to records, tapes, and discs. They come to view exhibitions or listen to lectures. Neither of these shows in the borrowing statistics. The number of borrowed books is not an exhaustive indicator of library output.

Another intricacy involved in productivity evaluation is that qualities are overlooked. Books differ in quality. How can this be measured in productivity assessment? There are other criticisms. The most important suggests that productivity as a measure of outputs does not capture what we really want to disentangle, namely the effects that the intervention, its implementation and outputs have produced with the end-receivers or in society at large; what are the value of these effects, and are the benefits worth the costs? In the library example, borrowed books are not significant by themselves; people may

charge out books from the library, place them in a heap on the desk at home, and after some weeks return them unopened. More important is the reading of the borrowed books. But what really matters are the borrowers' recreational or educational gains from their reading.

In sum, productivity as a yardstick of output is not an ideal measuring rod for assessing the worth of public sector activities. The public institution may do wrong things, i.e., the outputs may not produce the desired outcome. When used on a regular basis as a criterion of good performance, productivity might actually twist the organization away from its basic purpose. It may lead to goal-displacement, i.e., the organization will focus on outputs to the detriment of producing good outcomes.

Economic Models: Efficiency

While productivity is measured in the output stage, the second major economic model, efficiency, uses measures taken in the outcome stage. Efficiency is expressed as the ratio of the value of the outcome produced by the intervention to intervention costs. For instance, number of satisfied service users treated per man-year or simply outcome:costs. Or, alternatively, value of outcome created per euro spent.

Efficiency can be measured in two ways, as cost-benefit or as cost-effectiveness. If measured in a cost-benefit analysis, efficiency can be expressed as the ratio of the monetarized value of the outcomes produced by the program to the monetarized costs. If equalized to what is measured in a cost-effectiveness analysis, efficiency pays heed to monetarized costs as in cost-efficiency analysis, but the value of the effects is indicated in physical terms only. This is expressed in the simple algorithms in figure 1.7.

1) $$\text{value of program effect (in € etc.)} = \frac{\text{Efficiency (cost benefit)}}{\text{costs (in € etc.)}}$$

2) $$\text{program effects in physical terms} = \frac{\text{Efficiency (cost-effectiveness)}}{\text{costs (in € etc.)}}$$

Figure 1. 7 Efficiency as (1) Cost-Benefit and (2) Cost-Effectiveness

'Program effects' in figure 8 indicate consequences produced by the program. What we are looking for in efficiency analysis are consequences produced by the program and nothing else.

As promised from the outset, I will now end with a few words on one of the many possible procedural models that might be used in evaluation, namely deliberative democracy.

A Procedural Evaluation Model: Deliberative Democracy

Basic to the theory of deliberative democracy is the belief that representative democracy, identified with competitive elections of political leadership for legislative and executive offices, needs to be supplemented by more citizen involvement. Wide and intense participation in evaluations organized as deliberative processes might become such a supplement. Deliberation is an approach to evaluation in which particularly users—addressees, ordinary people, the man in the street, the common man—but also other intervention stakeholders consider relevant facts from multiple points of view, converse with one another to think critically about options before them and enlarge their perspectives, opinions, and understandings.

Supporters of deliberative democracy argue that all kinds of public sector evaluation might be organized in ways that encourage intervention users, other intervention stakeholders, and evaluators to reason with each other about intervention implementation, outputs and outcomes and how they should be valued in order to promote better decisions, hold power-wielders accountable, create a richer civic culture and a vibrating democratic polity.[7] Evaluations should be used as arenas within which more sustained popular participation and reasoned deliberation becomes possible. Instead of 'deliberative democracy' expressions such as 'dialogic democracy' (Giddens 1994:15-17 passim) and even 'discursive democracy' are used (Dryzek 1990).

The Oxford English Dictionary indicates the following meanings of the word 'deliberation':

1. The action of … weighing a thing in the mind; careful consideration with a view to decision.

2. The consideration and discussion of the reasons for and against a measure by a number of councillors (e.g. in a legislative assembly).

To deliberate is to carefully scrutinize and weigh up the pros and cons of statements of fact and the value and measures to be adopted with a view to making decisions. The International Encyclopaedia of the Social & Behavioural Sciences expresses the deliberative idea in the following fashion:

> 'In its political signification deliberation refers to a process of reasoning about political choices. In a deliberative process the actors are open to the facts, arguments, and proposals that come to their attention and share a willingness to learn from their colleagues and others… After a lengthy period of neglect by social scientists.., there was a resurgence of [the] study of deliberation in the last decades of the [20th] century. This new scholarship addresses both deliberation within governing institutions and, more extensively, deliberation among the citizens of a democracy.'

[7] 'Deliberative democracy' i *Dictionary of the Social Sciences,* Craig Calhoun, ed., Oxford University Press 2002.

Intelligent deliberation may occur among politicians in representative political assemblies. Yet, in the present context, the focus is set on deliberations outside the formal institutions of government in evaluations organized as deliberative arenas with many participants.

The criteria for an ideal deliberative discourse include:

1. Inclusiveness. Deliberation should be open to all competent speakers. No one with the competence to speak or act should be excluded. According to a radical interpretation: the deliberation process should directly include all those affected. According to a moderate interpretation: arguments put forward should be representative of all the arguments that those affected would or could have put forward.

2. Autonomy. Actors should be able to introduce the arguments they want and challenge those put forward by their communicative counterparts.

3. Power neutrality. No participant should be prevented, by internal or external coercion, from the rights laid down in Inclusiveness and Autonomy; no force except the force of better argument is allowed.

4. Ideal role taking. Actors should adopt an attitude of reciprocal respect. Actors must justify their positions to each other, listen to each other and respond to each other. They should not talk only, but also listen. All strategic action should be excluded. That is, actors should not view each other as a means, as limiting or enabling factors for the achievement of some other end than contributing to the substantive issue under deliberation.

These criteria testify to the fact that proponents of deliberative democracy get their inspiration from Jürgen Habermas's works on 'the ideal speech situation'.

Deliberation may produce not only changes in participants' original values and conceptions of the real world, it may also result in qualitatively more sophisticated preferences and more well-reasoned statements of reality.

Quality improvement of political views may take four different forms. Increased sophistication is one form. Opinions and world views of the participants may become more reasoned and much better argued than before. Reduced recourse to self-interested opinions is another upside. In the wake of rounds of deliberation participants tend to care less about their own personal interests and more about the public good. Enhanced tolerance for other people's very different points of view is a third advantage. A fourth upside is increased respect for the predicaments of political decision-making and political officials, for instance by acquiring increased confidence in elected officials and their work.

Quality improvement of – in deliberative democracy, decision-making is supposed to be different from the dominant mode in representative democracy. Ideally participants should listen carefully to each others' arguments

and positions and generate group decisions only after due consideration. Concerning the nature of consensus behind decisions there are several schools of thought in deliberative democratic theory. One school maintains that deliberation should aim at finding the objectively best solution because there are such things in the real world. According to a second, more pragmatic view, the point with deliberation is still to reach a reasoned consensus, although this is not necessarily the objectively best solution. According to a third position, dialogic democracy is not unavoidably oriented to the achieving of consensus. It does not imply that all divisions or conflicts can be overcome through dialogue, far from it. There are issues that remain essentially contested, even after much deliberation. Dialogic democracy presumes only that dialogue in a public space provides a means of living along with others in a relation of mutual tolerance. Deliberation and dialogue should be understood as the capability to create active trust through an appreciation of the integrity of the others (Giddens 1994:115 f.)

Deliberative democracy is justified in two main ways. It produces better decisions and better citizens.

The better-decisions and better, more rational and more inclusive decisions on the subject-matter toward which the deliberation is directed. Deliberation will support the development of new ideas, service concepts, solutions, and technologies in those subject areas toward which it is addressed. It will generate more fair and equitable outcomes as well.

The better – citizens argument suggests that wholehearted participation in deliberations will educate the participants on civic virtues so that they will become better members of the community in the future. This democratic schooling is expected to occur as a side-effect. Deliberations will ameliorate the participants' moral reasoning, cognitive faculties, and feeling for the common good. The participants will acquire civic skills which they may use later on and in other contexts. The effects may show already at the end of the current deliberation on the subject matter at hand. First and foremost, however, these effects are assumed to crop up in upcoming deliberations beyond the current one. It may show up in the next round of decision-making on matters that will later be placed on the political or administrative agenda. Deliberative democracy aims at the democratization of democracy. The demand for more deliberation in the carrying out of evaluations is regarded by some theorists as a 'democratizing of democracy' (Giddens 1994:16).

The better citizens argument regards deliberation as a vehicle for attaining something beyond good results in the current subject area. An evaluation organized in this fashion is an obvious intervention in itself. It is a vehicle for the enhancement of civic virtues. Such an evaluation may also be characterized as a carrier-rocket, because the deliberations in the evaluation group are only a carrier for the real rocket that is schooling for democratic virtues.

Consequently, if deliberative democracy is taken as the overarching merit criterion for an evaluation, the evaluation team should consist of a number of clients and other stakeholders in the intervention to be evaluated. There might also be a steering group of trained evaluators of course. These people should constitute themselves as a social community in order to deliberate the implementation, outputs and outcomes, side-effects included, of the in-

tervention. They should assess what has happened and how this ought to be judged.[8]

Final Note on Choice Among Evaluation Models

In this chapter, I have provided a broad overview of the evaluation landscape from a narrow perspective. I have organized the survey of the field by merit criteria used in the various general evaluation approaches.

Evaluation in general has developed from unity to pluralism. The total agreement which once existed in the early North American, Continental European, and Nordic evaluation communities on the appropriateness of the goal-attainment model has been replaced by a situation where multiple models compete.

For a long time, there was a strong tendency in evaluation literature to debase particularly the goal-attainment model, and its cousin side-effects evaluation. In recent years though, it has made a comeback in some sectors, partly under the label of effectiveness evaluation, partly as evidence-based social work, evidence-based social medicine, and so on. Yet criticisms are still dominant. On one significant account, I take exception to the criticism. From a representative democratic point of view the goal-attainment model and particularly the side-effects model are very important, since they are based on the conception of the parliamentary chain of influence. Intervention goals set by popular assemblies, governments, and agencies are not just any goals whatever. They are established by the constitutionally rightful representatives of the people or by the organs to which the representatives have delegated their power. Citizens, elected officials, and other principals have legitimate reasons to ascertain whether intervention goals have in fact materialized in the field. Otherwise, they cannot function as principals in the representative system of government.

The argument from representative democracy in favour of the goal achievement model, however, cannot remedy the fact that this model runs into difficulties particularly with goal catalogues. A major drawback is its lack of focus on side-effects. For this reason, side-effects evaluation is preferable to goal achievement evaluation.

Since the 1980s, economic models have gained ground (due to the deterioration in public finances and the neoliberal ideological wave that pushed the dissemination of the so-called New Public Management). Also these models can be justified by reference to the parliamentary chain of control and representative democracy. Citizens, their elected representatives, and their appointed executives need not only output and outcome information but also cost information. This is the strength of economic models. Another strength of the models are their ability to reduce the worth of an activity into a plain and simple number, easy to grasp and remember. It must be kept in mind, however, that like other designs, they provide partial perspectives only. The danger with economic models is that decision-makers are fascinated by their mathematical precision and wrongly believe that they provide comprehensive, final answers.

[8] Jon Elster (1998:153 ff.) delivers an array of insightful objections to this theory.

Another tendency which started already in the 1970s is the use of professional models. They are applied particularly in fields dominated by complicated, informal criteria of merit and strong professions. Higher education and academic research are paradigm cases. Professional models differ very much from goal-based models and economic models in taking the quality criteria of the pertinent profession as their evaluative point of departure. These criteria are customarily tacit and unwritten. Professional evaluation mostly proceeds as evaluatee self-evaluation followed by external peer review by independent, highly qualified colleagues. The use of statistical and other data, written documentation but also evaluator-evaluatee dialogic interaction and deliberation are important features of professional evaluation. The strength of the professional models rests with their ability to capture and judge qualities.

Due to the increasing criticism of the public sector and the representative system of government, the demand for a larger participation by clients and other stakeholders has increased. This has resulted in an intensified use of client and stakeholder models. Stakeholder models provide the broadest view possible of government interventions, and promise to take the views of all involved into consideration. They proceed by in-depth interviews and stakeholder deliberation. They usually entail an interesting combination of social research and political accommodation of various stakeholder interests in order to shape social agreement and render legitimacy to decisions. On the downside their impracticality should be mentioned. They are very time-consuming and demanding to handle.

Client-centred models have an increasingly important role to play, particularly in government social services, but also in parks, recreation, cultural activities and education. Client-oriented evaluation is probably partly based upon the idea of citizens as customers. Yet the basic notion behind at least some of them comes from the doctrines of participative and deliberative democracy. Citizens should not participate in public sector activities as voters only. They should also participate as users and other stakeholder role incumbents. User-orientation might breed effectiveness, legitimacy, and service accommodation to the participants. But client-oriented evaluation and stakeholder evaluation should be kept inside the frameworks instituted by the representative democratic system. In this fashion stakeholder and client evaluation may supplement goal-based models and economic models.

The deliberative democracy approach to evaluation is gaining ground in evaluation discourse and practice. Preferably, it should be used in combination with client-oriented, stakeholder-oriented, or profession-oriented models. The basic idea is that the process of collecting and analysing information on intervention implementation, outputs and outcomes and appraisal of intervention merits should be carried out dialogically and interactively between evaluators and evaluatees and among evaluatees internally. Such a procedure would promote not only more informed views on interventions but also, as a side-effect, produce better citizens.

The danger with all evaluation models is that they might be applied too uncritically and that decision-makers wrongly believe that one model can provide comprehensive, final answers. Therefore, it is important to keep in mind that every model provides only partial perspectives and answers. For this reason, combinations of several models are commended. At present,

pluralism, not uniformity, is the most reasonable attitude towards the proliferation of value criteria characterizing latter-day evaluation.

At least some of the evaluation models outlined in the present chapter are partial also from a procedural perspective. They must be supplemented by other value perspectives normally demanded from public sector activities in contemporary democracies: legal equity, procedural fairness, class, gender, and ethnic representativeness, participatory values, and publicity rules. No evaluation, however pretentious from a scientific and scholarly point of view, can explain in an objective fashion how a balance should be struck between these values and substance-only and economic values. The trade-off can only be made through public debate, opinion formation, intensive deliberations, compromise, and eventual majority decisions, i.e. through politics and policy-making.

It is my hope that experienced researchers, practitioners, and other actors in social work and social welfare will get inspiration and eye-opening flashes from my general exposition so that they can develop what is valuable and disregard what is inappropriate.

References

'Deliberative democracy' in Dictionary of the Social Sciences, Craig Calhoun, ed., Oxford University Press 2002. Oxford Reference Online. Oxford University Press. Uppsala Universitet. 17 December 2003 http://www.oxfordreference.com/views/ENTRY.html?subview=Main&entry=t104.e428

Elster, John, 1998, ed., Deliberative democracy, Cambridge: Cambridge University Press.

Giddens, Anthony, 1994, Beyond Left and Right: the Future of Radical Politics, Oxford: Polity Press.

Vedung, Evert, 2000, Public Policy and Program Evaluation, New Brunswick, New Jersey: Transaction, paperback, www.transactionpub.com/-cgi_bin/transactionpublishers.storefront.

Vedung, Evert, 2002, 'Utvärderingsmodeller' (Evaluation Models), Socialvetenskaplig Tidskrift (Special issue on evaluation and social work), 9:2-3, 118-143. Also published by Uppsala University, Institute for Housing and Urban Research, 2002, ISBN 91-506-1643-9.

Organising projects in a welfare context - challenges to evaluation

Stefan Sjöblom

Introduction

There are many indications that projects as a form of organisation and management have increased especially during the past decades. As Rehn (1999) argues, the project has become a post modern symbol of adaptability and contingency – it is thought of as the superior way of reacting to unforeseen and non-standard situations. The phenomenon is apparent both in business and public administration. According to a French study, the use of the term project started to increase in the 1960s (Boutinet 1996, also Lundin & Midler 1998, 234). This observation is well in line with the institutionalisation of project management. The world wide association IPMA (International Project Management Association) was founded in 1965. At present the American association PMI (Project Management Institute), founded in 1969, has approximately half a million members and the number is increasing by 25% annually.

Since the 1960s project management has become increasingly established as a management field in its own right (Blomquist & Söderholm 2002, 25). Apart from professional associations, places within university curricula and certification procedures have also emerged. the PMI lists 24 universities in the USA granting graduate degrees in project management and 19 outside the USA (Anell & Wilson 2002, 183). Project management standards have, since 1987, been published and circulated worldwide by the PMI in the so called PMBOK (Project Management Body of Knowledge). During the past five years, ISO standards for project management have been developed, including terminology standardisation (ISO 15188:2001), as well as quality management (ISO 10006:1997). In a Nordic context, the Nordnet was founded in 1981 as an informal forum for cooperation between Nordic project management associations. Today such associations exist in all Nordic countries.

As these examples show, project management has been especially anchored within the fields of business administration and engineering. In the 1960s and 1970s the search for the efficient project was part of the rationalistic dream (Engwall 1995). The instruments and steering techniques started to emerge in weapon systems development in the USA during the latter part of the Second World War and during the Cold War era (Blomquist & Söderholm 2002, 27). Perhaps the most famous method of all, PERT – Program Evaluation and Review Technique – was developed in 1957 to control and evaluate large development projects. PERT has had several successors that have been developed into commercial products by companies throughout the world—in the Scandinavian context for instance by Eriksson and ABB (Blomquist & Söderholm 2002, 27 f.). By and large, project management has in various ways been able to create a strong, normative pressure and the rationalistic approach to project management is still very much present.

This rationalistic bias may today seem somewhat strange, as research has shown that projects do fail. Project failures are for instance caused by unclear goals or deficiencies in management, such as insufficient planning or lack of co-ordination and consideration for environmental/political factors—such as opposition by important stakeholders (Engwall 2002, 261). Since such failures are not entirely unknown in public decision-making either, the lack of systematic information and scientific discussion on the increase of project administration in the public sphere is surprising. It should however be said that important work as been done by the Scandinavian School of Project Management in developing an understanding of projects by combining an organisation theory perspective with theories on entrepreneurship and industrial development (Sahlin-Andersson & Söderholm 2002, 11). Although the focus of the School is on industrial projects, there is a clear ambition to go beyond the business world by analysing the implications of 'projectification' for society in general (cf. Lundin & Söderholm 1995, Lundin 1998, Sahlin-Andersson & Söderholm 2002). Political scientists and sociologists have, however, only to a very limited extent been active parties in this discussion.

Although the attractiveness of projects in the private and the public sphere may rest on a fairly homogenous set of expectations such as controllability, flexibility and change (Sahlin-Andersson 2002, 241), there is a lack of theoretical conceptualisation of project administration within the public sphere. The 'project state' can, due to different contextual contingencies, operate only to a limited extent according to the same logic as the industrial company organised according to project management principles. Within the public sphere the project organisation is one example of the blurring boundaries between various modes of organising and an increasing involvement of competing interests. Projects emerge and operate in relation to the politico-administrative structure. This means that projects develop in a variety of inter-organisational settings.

Although the importance of evaluation has grown immensely in the public sector in general during the past decades, partly as a reaction towards the rationalistic dream of the 1960s and 1970s (cf. Vedung 1997), evaluation is today especially related to program and project management. The European Commission, for instance, includes the use of evaluation as a key element in improving the management culture of the Commission itself (European Commission1997, 7). The so-called SEM 2000 (Sound and Efficient Management 2000 initiative) included the requirement that systematic evaluation should be introduced for all EU programmes. Partly because of the European Union and partly for domestic reasons, project proliferation seems to be a predominant feature of the public sector in many European countries. The main question in this article is to what extent the specific features of project organisation constitute challenges to evaluation criteria and evaluation designs. A strong presumption behind the article is that project organisation as a phenomenon is growing more and more complex to the extent that the challenges are numerous.

Project organisation – definitions and conceptions

Empirical ambiguities

An assessment of the general importance of project activities within the public sector is hard to make. The reasons are empirical as well as conceptual. Empirically speaking, projects are hard to count (cf. Sahlin-Andersson Söderholm 2002, 11 f.). There are for instance no comprehensive studies on the extent or number of projects or project-type organisations in the Finnish public administration. Available data are limited and incomplete. There are however indications that the phenomenon has been increasing. The project register maintained by Ministry of Finance indicated that about 700 projects were established at the ministerial level in year 2003, 300 of which were defined as development or renewal projects (http://www.hare.vn.fi). The same figures for the year 2000 were 450 and 260 respectively. Moreover, there are thousands of structural fund projects at the regional level. The number of locally initiated projects is of course impossible to estimate.

Anell & Wilson (2002, 171) suggest that present organisational life is characterised by two tendencies. On one hand routine-based organisations appear to be growing more projectified. On the other hand, temporary organisations are becoming more routinized, taking on characteristics from routine-based organisations (Anell & Wilson 2002, 172). The two tendencies indicate that an increasing projectification may on the one hand reflect the ambition of permanent and routine-based organisations to answer to demands of innovation and change. On the other hand, it may reflect efforts of organisations working in unstable environments to routinize the temporary features of the organisation in order to cope with the uncertain and unforeseeable features of 'adhocracy'. Presumably both tendencies mean an increasing visibility of project organisation in society.

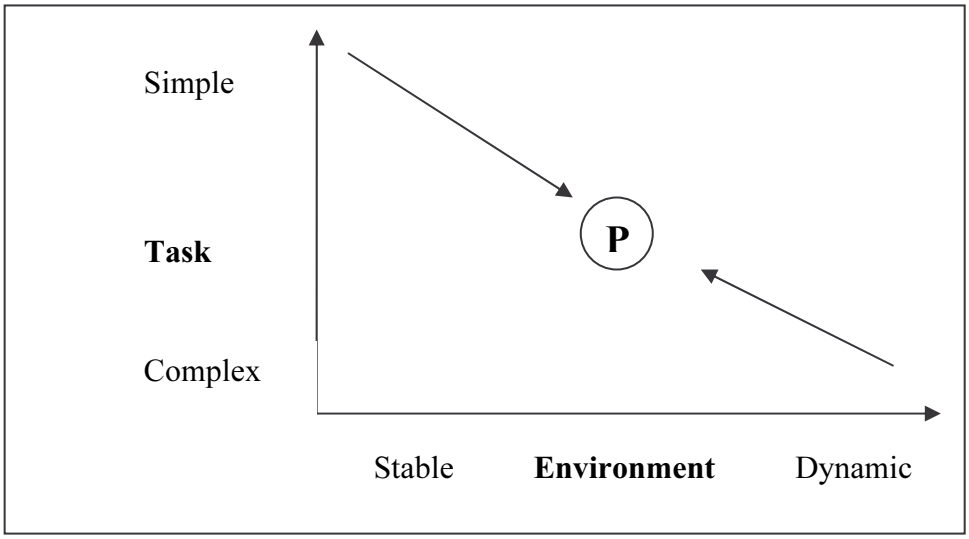

Figure 2. 1 A projectified society – two tendencies

Conceptual ambiguities

In conceptual terms the definitions of a project organisation differ. The PMI defines a project as ”a temporary endeavour undertaken to create a unique product or service” (PMI 1996, 4; also Engwall 1998, 25). This is an example of instrumental and even normative definitions of projects. The funding principles of the European Union as well as the conception of partnership have strengthened the inclination to define almost all reform activities as projects. According to the EU, a project is ”a single, non-divisible intervention with a fixed time schedule and dedicated budget” (European Commission 1997, 14). Similar project characteristics frequently referred to in the project management literature are (cf. Packendorff 1995, 320, Lundin 1998,197):

- a unique, once-in-a-lifetime task
- a predetermined time frame
- subject to one or several performance goals (such as resource usage and quality)
- a number of complex and/or interdependent activities

Such criteria may have a pedagogical point in drawing attention to some ideal features of project organisation. Analytically speaking they are however not particularly helpful. Project organisation in the public sector can rarely be isolated in terms of such characteristics. To an increasing extent tasks that are by no means unique are organised as projects. The time frame as well as the performance goals may be extremely ambiguous. Projects are more or less embedded in surrounding organisational structures and the contextual relationships may vary enormously. Projects do not emerge in a vacuum. They emerge in interaction with one or several permanent organisations or other temporary organisations and they are expected to provide these organisations with some kind of surplus compared to traditional forms of organising. As shown in figure 2, the project may, once the tasks are accomplished, remain an isolated (unique) case, it may generate new projects, or the tasks may at some stage be obtained by the line organisation. The surplus, whatever it is, is generated by disengaging the tasks in question from the line organisation.

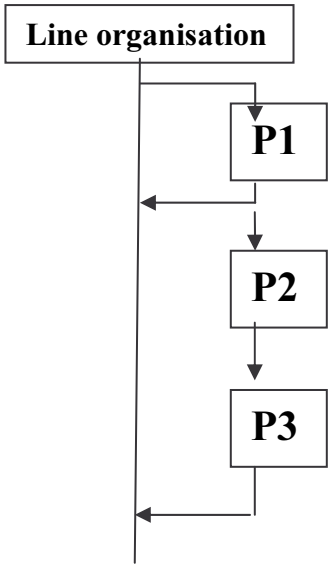

Figure 2. 2 The permanent organisation (line organisation) and the project

A project can be initiated as a reaction to various types of incentives inside or outside the administrative structure. The incentive can be a policy problem but a project may also be developed - for instance – as a result of diffusion or as a purely symbolic action. Regardless of reason, the project organisation is presumably developed through negotiations between the administration and other affected interests within the policy field. The composition of the project organisation and the nature of the policy problem determine the features of the decision-making procedures and the activities of the project can be pursued in more or less close relationship with the permanent administrative structure.

In more general terms, two conceptions of project organisation is often referred to in the project management literature: an instrumental view and a conception of projects as temporary learning organisations (Packendorff 1995, 325 ff.). According to the instrumental view, a project is a tool for attaining organisational ends. From that perspective a project consists of distinct time-limited sequences from development via implementation to termination. (Packendorff 1995, 321). Goal attainment is the primary criteria for evaluating projects and the goal attainment model is probably still the most frequently used model in project evaluation. The instrumental view has many limitations, one of which is that projects can be initiated for very unclear reasons. Moreover, the various motives for individuals participating in project organisation is also neglected (Packendorff 1995, 326). When projects are regarded as temporary organisations, the learning processes are emphasised rather than the goals. What is to be studied is temporary organising processes "i.e. the deliberate social interaction occurring between people working together to accomplish a certain, inter-subjectively determined task" (Packendorff 1995 328). The deliberative features of the processes are supposed to promote innovation and change. Although the two perspectives are not necessarily contradictory ones, especially the temporary organisation concept points to the facts that the various expectations, motives and values pursued through project organisation may be extremely varying and also

subject to constant change during the process (figure 2.3). The extreme variation in project organisation is easy to illustrate in terms of some recent public sector developments.

Phase:	Projects as tools	Projects as learning processes/ temporary organisations
Development	Goal	Expectations
Implementation	Choice	Action
Termination	Effects	Learning
Primary focus of evaluation:	Goal attainment	Organisational learning Innovation

Figure 2. 3 Project as a phenomenon – two conceptions

(based on Packendorff 1995, 328)

Why do projects emerge?

Generally speaking, the organising of projects is frequently regarded as a means of breaking with stability and bureaucracy in order to handle increasing and often inconsistent demands. They are expected to provide flexibility. In empirical terms it is clear however that the project as a phenomenon appears very differently depending on how temporary and permanent features of the organisations are interlinked. There are also considerable variations depending on the administrative levels studied. Moreover, a project organisation can be initiated to promote several administrative values, a fact which should affect the evaluation design and the criteria used in evaluations. Referring to the administrative development during the past decade it seems that most important administrative values pursued through temporary organisations have been legitimacy, efficiency/effectiveness and innovation. A preliminary systematisation of types of projects based on these two dimensions is outlined in figure 2.4.

Administrative value: / Administrative level:	Legitimacy	Efficiency/ Effectiveness	Innovation
Central state level	I. Projects as arenas for legitimising and regenerating the central state	III. Projects as effects of national steering systems (result orientation)	V. Projects as arenas for national and private research and innovation policies
Regional and local level	II. Projects as arenas for legitimising EU-policies in the member states	IV. Projects as arenas for coping with economic constraints	VI. Regional development and innovation projects

Figure 2. 4 A preliminary systematisation of various types of project organisation in the public sector.

The categories are by no means pure ones. Several features and values might be attached to the same project. In the absence of systematic empirical evidence the systematisation is also partly impressionistic, but it illustrates the great variety of project organisation in the public sector. The categories can be briefly described as follows:

A predominant feature at central state level seems to be that project organisation is increasing "on top" of administrative routines. During the 1990s there has been a development from conventional administration via New Public Management to governance. In the case of Finland, for instance, the traditionally most important arena for pursuing administrative reforms has been the committee system, including both permanent statutory committees and temporary governmental commissions (Sjöblom 1994). During recent years there have been significant changes from committees to ad hoc working groups with regard to the reform policy as well as the policy preparation in the ministries in general (Sjöblom 2000, Temmes 2001). The rising interdependence of the ministries led to lack of co-ordination at the level of central governmental (Bouckaert, Ormond & Peters 2000). One interpretation of this development would be that we are witnessing a renaissance of planning and co-ordination, to a considerable extent organised by means of ad hoc projects (Temmes 2004). An illustration of this can be seen in the cross sectoral strategic steering programs of the present government. Project management in connection to state reform policies can be regarded as an arena for legitimising and regenerating the central state (I.).

Empirically speaking, the field most characterised by projects is without doubt the regional development administration. The financial support from the EU to the member states is allocated by means of developmental programmes and projects, which on the one hand motivate an intense involvement in the negotiations and project work by the actors involved, and on the other, make it more likely for the impact of the projects to be extensive at all levels. An interesting interpretation would be that project organisation not only serves as a mechanism for financial distribution but also as a legitimacy-raising mechanism (II) at the output side of the policy cycle (Vedung

2004). One reason for this is that the EU as an administrative structure has weak legitimacy on the input side of the process (nationally elected representatives and nationally appointed Commissioners). The partnership approach is one example of the legitimacy-raising functions attached to project organisation. It has enabled a far broader spectrum of actors to become active in the decision-making processes, thereby broadening the scope for strategic development (cf. Marsden & Franklin 2004). Project organisation has an important role in 'trickling down' global regulatory frameworks to locally implemented initiatives.

Although there is a lack of systematic empirical data, it seems moreover likely that steering systems such as the strong result orientation in the public sector as a whole have caused project proliferation (III). Projects are regarded as delimited and efficient.

Since the economic crisis in the beginning of the 1990s, the public sector has suffered from severe economic constraints and under-budgeting. Project organisation at all administrative levels is at least to some extent serving as compensating economic mechanisms; it is a means of coping with economic constraints (IV). Beside the innovative and synergetic objectives usually attached to them, they are also used to obtain additional funding. Especially at local level, staff have been transferred to projects in order to reduce the personnel costs of the permanent municipal organisation. However, there is little systematic empirical evidence as to what extent this is the case. Furthermore, at the central state level in Finland there are indications of a continuously increasing 'projectification' due to under-budgeting, especially within welfare and healthcare services (Sulkunen 2004).

Finally, the so-called innovation projects have different implications depending on the administrative level concerned. The central State is of course also an important financer of research and development projects within the private sphere. Thereby, state actors are increasingly involved in the development and innovation strategies of the private sector (V). In terms of regional development within development and innovation projects (VI), the metaphor of partnership is the most significant symbol of the cooperation and interplay between the public sector, the 'third' sector and the private sector. It is to a large extent pursued through a network approach, which is expected to provide advantages in terms of synergy as well as innovation. The networks are supposed to develop into deliberative arenas organised as projects (Vedung 2004). They are not only an instrument for implementing EU- or nationally initiated programs or reforms, but also a new form of policy making – including lobbying.

Challenges to evaluations

As the examples in the previous chapter show, project organisation may be a means for flexibility in coping with inconsistent demands, but it is also introduced to promote several administrative values, such as legitimacy, effectiveness and innovation. In EU-programmes, sustainability as well as synergetic effects in terms of the partnership principle is often emphasised. By and large, the values and expectations attached to projects may be manifold as well as ambiguous. It seems clear that neither an instrumental nor a learn-

ing approach is necessarily sufficient in conceptualising project organisation in the public sphere. Such conceptualisations have to focus on the processes as well as the contextual features related to project organisation.

With regard to evaluation designs, however, there does not seem to be any reason for an a priori emphasising or abandoning of any particular evaluation approach. Projects as well as programs and policy instruments can be evaluated according to several approaches and models. Although the evaluation imperative today is strongly related to project management, there is still a great variation in the extent to which public projects are evaluated and the quality of such evaluations. When evaluations are undertaken, formative and summative evaluations are still the most commonly used designs.

In summative evaluations there appear to be very strong arguments in favour of multiple criteria designs as well as side-effect evaluations (cf. Vedung 1997). Side-effect evaluations offer possibilities for dealing with complexity and uncertainty in policy problems. Multiple criteria give a broader understanding of the effects of the project. However, several criteria can be used for many different purposes, a fact which puts strong demands on available data (cf. Hilden & al. 2002, 122). The flexibility of a regional development project, for instance, can be understood in terms of adjusting outputs to local needs and conditions, but the project can also provide the possibility of adjusting local expectations and preferences to match national or transnational regional development policies.

The ambiguity of project goals and expectations indicates that strong arguments can be made for a third approach, namely planning evaluation. Such arguments have been made for instance by Rossi & Freeman (1993), and also by the American National Science Foundation (1994). Planning evaluation is supposed to provide a "rich, context laden description of a project, including its major goals and objectives, activities, participants and other major stakeholders , resources, timelines, locale and intended accomplishments" (National Science Foundation 1994, 4). By and large the purpose of planning evaluation is to assess understanding of project goals, strategies and timelines.

Although there may be many problems associated with planning evaluations, the point is that coping with ambiguity means understanding the initial phases of the project and the ways in which the interpretation of goals, tasks and strategies change during the process. This has to do partly with a phenomenon that Kristian Kreiner calls the problem of dual realities (Kreiner 1996, 339 f.). Although the members of the project organisation may get a continuous response from various stakeholders, they interpret the information according to their own expectations. The phenomenon has been regarded as an explanation for the fact that project organisation may persist long after the stakeholders have abandoned the project as irrelevant. The tacit knowledge of the project members may have considerable effects on the development of the project.

Emphasising the ambiguity and the process related features of projects is important because of the previously mentioned lack of conceptualisation of the possible effects of project organisation on the temporal as well as the structural context of public decision-making. Lundin & Söderholm (1995) argue in favour of an action-based theory on temporary organisations. One

particular purpose of such a theory would, as they say, “be to account for mechanisms guiding the apparent action orientation in society today” (Lundin & Söderholm 1995). Project organisation represents one such mechanism. We basically agree on this point of departure. Important dimensions in such a conceptualisation are at least (cf. Lundin & Söderholm 1995, 439) time, task (policy design in terms of the relationship between projects and policy programmes), the composition of the project organisation, and transition, i.e. the qualitative difference before and after the accomplishments of the project organisation. All dimensions are crucial, especially as the element of change is strongly related to project organisation. They are perceived to be created in order to fulfil a special purpose. But they may also, along with many other organisational mechanisms, be used in order to prevent changes (Brunsson & Olsen 1990). In all these respects project evaluation has an important role to play in broadening our understanding of the relationship between temporary and permanent features of public decision-making.

References

Anell, Barbro I – Wilson, Timothy L. (2002): Organizing in two modes. In: Sahlin-Andersson, Kerstin – Söderholm, Anders (eds.): Beyond project management. New perspectives on the temporary – permanent dilemma. Liber Ab. Malmö.

Blomquist, Tomas – Söderholm, Anders (2002): How project management got carried away. In: Sahlin-Andersson, Kerstin – Söderholm, Anders (eds.): Beyond project management. New perspectives on the temporary – permanent dilemma. Liber Ab. Malmö.

Bouckaert, Geert, Ormond, Derry and Peters, Guy: A Potential Governance Agenda for Finland, Ministry of Finance, Research Reports 8/2000, Helsinki 2000.

Boutinet, J-P (1996): Anthropologie du projet. Presses Universitaires de France. Paris.

Brunsson, Nils – Olsen Johan P. (1990): Makten att reformera. Calssons. Helsingborg.

Engwall, Mats (1995): Jakten på det effektiva projektet. Nerenius & Santérus. Stockholm.

Engwall, Mats (2002): The Futile Dream of a Perfect Goal. In: Sahlin-Andersson, Kerstin – Söderholm, Anders (eds.): Beyond project management. New perspectives on the temporary – permanent dilemma. Liber Ab. Malmö.

European Commission (1997): Evaluating EU Expenditure Programmes: A Guide. First Edition, January 1997.

Hilden, Mikael & al. (2002): Evaluation of environmental policy instruments. Monographs of the Boreal Environment research. Monograph No 21. Finnish Environment Institute. Helsinki.

Lundin, Rolf A (1998): Temporära organisationer – några perspektivbyten. i: Czarniawska, Barbara, red. (1998): Organisationsteori på svenska. Liber Ekonomi, Kalmar.

Lundin, Rolf & Söderholm, Anders (1995): A Theory of the Temporary Organization. Scandinavian Journal of Management Vol 11, No 4.

Lundin, Rolf A. & Midler, Christophe, eds. (1998): Projects as Arenas for Renewal and Learning Processes. Kluwer Academic Publishers, Boston.

Marsden, Terry – Franklin, Alex (2004): Projects in rural and regional development: tools for innovation or arenas for a "project class"? " In.: Andersson, Kjell – Eklund, Erland – Sjöblom, Stefan (eds.): Theory and Practice of Governance in the Project State. (forthcoming).

National Science Foundation (1994): User Friendly Handbook for Project Evaluation.

Packendorff, Johann (1995): Inquiring into the Temporary Organization: New Directions for Project Management Research. Scandinavian Journal of Management Vol 11, No 4.

Rossi, P.H. & Freeman, H.E. (1993): Evaluation – A Systematic Approach (5th edition). Sage. Newbury Park.

Sahlin-Andersson,Kerstin (2002): Project management as boundary work. In: Sahlin-Andersson, Kerstin – Söderholm, Anders (eds.): Beyond project

management. New perspectives on the temporary – permanent dilemma. Liber Ab. Malmö.

Sahlin-Andersson, Kerstin – Söderholm, Anders (eds.): Beyond project management. New perspectives on the temporary – permanent dilemma. Liber Ab. Malmö.

Sjöblom, Stefan (1994): Finland. förvaltningspolitiken 1972-1990. I Laegreid P & Pedersen O.K. (red.) Förvaltningspolitik i Norden. Jurist- og okonomforbundets Forlag. Köpenhamn .

Sjöblom, Stefan (2000): Hallinnon uudistusten valmistelu pohjoismaissa. Teoksessa Mäki-Lohiluoma, Kari-Pekka & Oulasvirta, Lasse (toim,) Tuleeko neuvottelemalla tulosta? : valtion ja kuntien neuvottelujärjestelmä. - Tampere : Tampere University Press.

Sjöblom, Stefan (2004): Towards a projectified society: On the logic of "the project state" In.: Andersson, Kjell – Eklund, Erland – Sjöblom, Stefan (eds.): Theory and Practice of Governance in the Project State. (forthcoming).

Sulkunen, Pekka (2004): The power structure in the project state: how "bottom up" became "top down" in public administration In: .: Andersson, Kjell – Eklund, Erland – Sjöblom, Stefan (eds.): Theory and Practice of Governance in the Project State. (forthcoming).

Temmes, Markku. Julkinen johtaminen (Public management), VAPK, Helsinki 2001.

Temmes, Markku (2004): Public administration reforms and project governance in Finland. In.: Andersson, Kjell – Eklund, Erland – Sjöblom, Stefan (eds.): Theory and Practice of Governance in the Project State. (forthcoming).

Vedung, Evert (1997): Public Policy and Program Evaluation. New Brunswick. Transaction Publishers.

Vedung, Evert (2004): Evaluation of project-oriented network management? . In.: Andersson, Kjell – Eklund, Erland – Sjöblom, Stefan (eds.): Theory and Practice of Governance in the Project State. (forthcoming).

2 PERSPECTIVES ON EVALUATION

Evaluating effects - utopia? [9]

Pekka Sulkunen

1. "What works?"

The expression what works originated in the area of crime prevention, particularly in prison programmes intended to rehabilitate convicts and prevent recidivism. A vivid debate ensued when Robert Martinson published his widely read and cited review of prison rehabilitation programmes with the title "What Works in Prison Reform?" (Garland 2001, 58). Martinson came to the conclusion that "nothing works", if the target is the reduction of recidivism and return to normal non-criminal life for the convicts. Despite this gloomy start of the cliché, it has since then been used in a wide range of treatment, prevention and promotion contexts, including social work, health promotion and treatments for different types of economic, social and psychological disorders, especially those related to exclusion.

The appeal of the expression is its blunt simplicity and the implied pragmatism. Its users imply that they are not interested in theories concerning the ethiology of the problems, or the origins or causes in the personal history of the clients, or even in the conditions that lead to the problem. It is a very professional question, but a question that practitioners like to ask rather than social scientists, psychologists or philosophers.

As such it further implies a limited focus on the function that the practitioners are responsible to fulfil. What works is a relevant question only if (1) there is a unity of goals and (2) there is a consensus concerning this unity, and (3) the attainment of these goals can be measured and the effects of the "what" can be ascertained.

In alcohol and drug prevention, especially if it is targeted at young people, these conditions are not normally fulfilled. In the normal case, the goals tend to be contradictory rather than a unity. Reducing risks or increasing risk awareness implies increased awareness of the substances and their effects, and this is likely to imply greater frequency of use. Reducing parents' anxieties by information and parenting-related skills are more than likely to increase other adults' anxieties, and so on. Secondly, in contemporary societies it is very difficult to reach a consensus about drug and alcohol issues, even in such basic questions as should teenagers be served or supplied with alcohol or not. Thirdly, the effects of almost any types of prevention programmes, especially at the community level, are likely to be small, confounded with other influencing factors than the programme itself, and if defined in a way that allows measurement, often rather uninteresting and transient. For example, we might be able to measure an element in drug-related

[9] This article is based on a keynote speech in the seminar: Targeted Drug Prevention How to Reach Young People in the Community? Pompidou Group, Helsinki 5.11.2002 (revised 14 March 2003). Published in the conference report: Targeted Drug Prevention. Conference Helsinki 5-7 Nov 2002.

risk awareness and ascertain that the programme was responsible for it, but such a limited goal is not of major interest in itself, if nothing changes in drug-taking practices, or if the drug-taking practices change in a way that are considered undesirable from the point of view of the programme theory.

"What works?" is also a question that implies that there are possibilities for accumulating evidence from evaluations. It is a question that leaves out the context, and suggests that standardised programmes could be developed for a wide variety of different types of cultures and situations. This also is usually not true. Even highly structured social skills training programmes or educational packages tend to be interpreted differently in different contexts and have quite different effects accordingly.

Are we then to conclude that the search for best practices and evidence-based policies is probably an error that can only lead to disillusionment, frustration and apathy at worst, or a professional fantasy that in the best of cases helps to maintain the integrity of practitioners without doing much harm, if not any good either?

From a sociological point of view such a conclusion would be hasty and as unfounded as the naive acceptance of the excessive claims of evidence-based approaches to prevention work.

2. Effect, impact, outcome and efficiency

To answer the second question of my title, "What can be based on evidence?" we must look at the case much closer. The term "evidence-based" in prevention or promotion contexts is also a loan, not from criminology this time but from medicine. The so-called Cochrane collaboration is intended to provide rigorously qualified meta-analyses of studies on the effects of specified medical techniques. The requirements for acceptance in the corpus of studies that can be included in the meta-analyses are very strict and strongly prioritise randomised control trials (RCT). The purpose is to eliminate or to minimise not only harmful practices but also costs of not well-founded treatments so that limited resources could be allocated efficiently to practices that actually are useful.

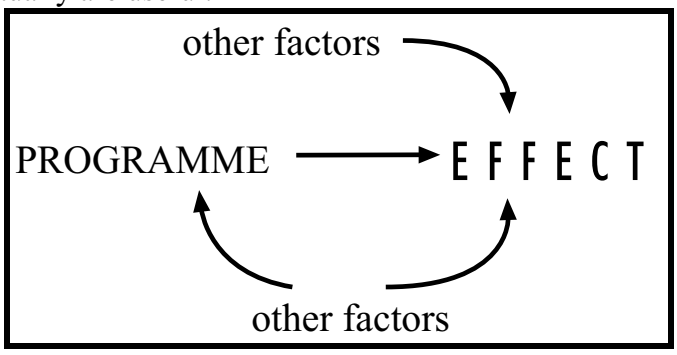

Figure 3. 1 Looking for effects

The Cochrane criteria for inclusion in fact only look at a limited type of evidence, namely those concerning the effects of a proposed treatment (Figure 3.1). Evaluation researchers have their own terminologies, and therefore

there is a need to explain what is meant by each of the terms. Looking for effects means in my terminology the research operations that are required for identifying a specific anticipated positive result from a treatment, or in our case, prevention programme, in a way that eliminates the context from the focus of interest. The RCT is the ideal arrangement for this type of research. Looking for impact in my terminology means that we are interested not only in the question whether the trial programme is likely to have the anticipated specific effect but more generally in the intended and unintended effects that the programme may have by generating a social process (Figure 3.2).

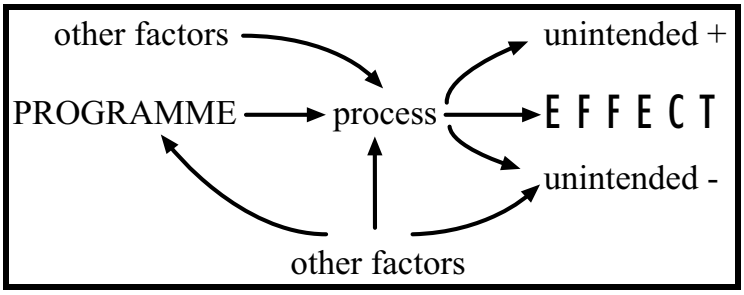

Figure 3. 2 Looking for impact

Looking for outcomes means that we are not only interested in the effects of the programme on the goals of the prevention programme, but more widely in the consequences of implementing it, not only as regards the programme goals but also on other things such as working motivation, multi-professional relationships, the role of the NGO's, issues of power etc (Figure 3.3).

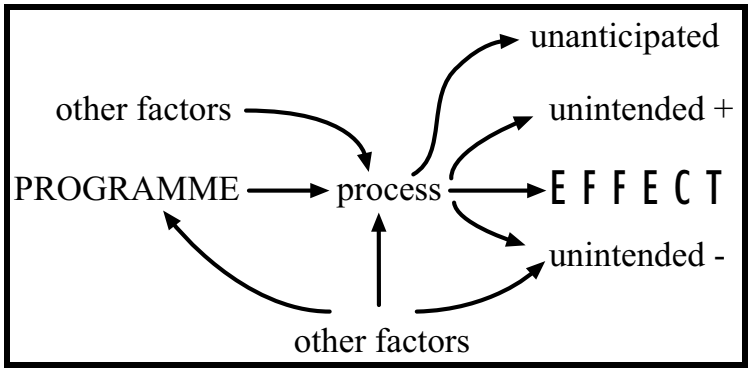

Figure 3. 3 Looking for outcomes

And finally, looking for efficiency means that we are interested not only of the consequences of the programme but also in the contextual factors that influence the feasibility, the meanings associated with the process, as well as in the costs involved.(Figure 3.4).

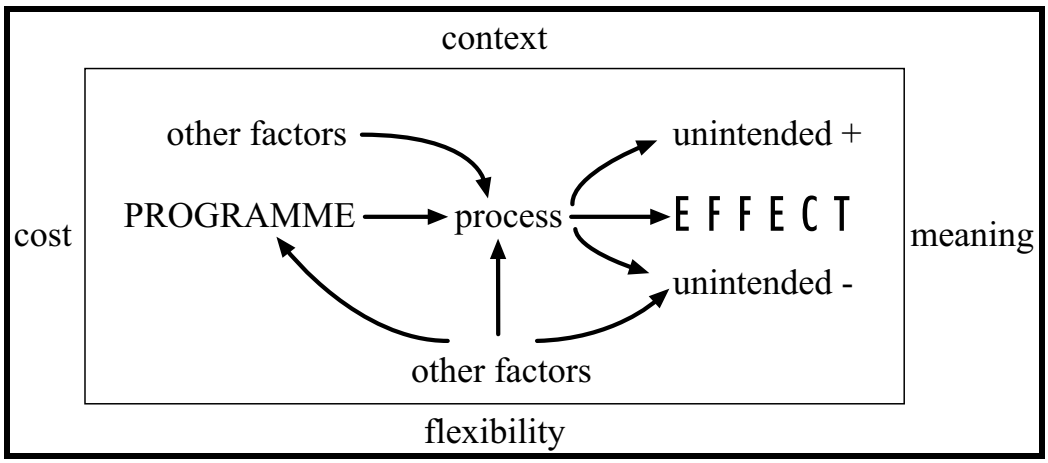

Figure 3. 4 4 Looking for efficiency

3. The importance of complex meta-analysis

From practical experience we know that results from effect-oriented programmes and their evaluations are seldom transportable. They are usually small, if the target is set at use practices, and at most modest if the target is public opinion or youth attitudes towards drugs or drinking. Studies that are powerful enough to detect such minor changes require huge samples to be statistically significant, ant therefore they are very expensive. For these reasons we actually have very few such studies available, and are very unlikely to have very many in the future. For this reason alone the Cochrane approach that has been approximated in so-called Campbell collaboration (www.campbellcollaboration.org) is hardly useful in the prevention context at all - why build up complicated networks to review and sum up those few available studies that everyone interested can pick up for themselves to read and judge?

Instead, we should build up international effort and synergy in looking systematically at the more complex evaluation studies on impact and outcome, and in special cases also on efficiency. There are thousands of prevention projects in the world that can teach us about why certain types of programmes and projects are popular in contemporary European contexts - for example of the type we studied in Helsinki (Rantala & Sulkunen, forthcoming). We pointed out that the central difficulty of establishing a community-based prevention programme in contemporary pluralistic society is the lack of moral resources on the part of the public authorities in discussing and advicing the population about the correct attitudes towards the multiplicity of drug use. This same difficulty appears in a number of other areas in health promotion/disease prevention. We can accept the expert evidence on risks, but have difficulty in deciding what individuals should do about them. A meta-analysis of similar studies would help us see in what ways programme managers have reacted to this problem, and would educate us on the kind of advice they need in these very typical situations.

Such meta-analyses can not be squeezed to simple formula about effects. In order to write them sensibly, one must be equipped with a sensitivity not

only to the specific issues related to drug prevention as such but to the characteristics of contemporary societies at large, to the social processes that can be initiated by public policy, and to the likely reactions of the populations to them. Therefore such meta-analyses are complex and require a high level of social science training.

References

Garland, David: The Culture of Control. Crime and Social Order in Contemporary Society. The University of Chicago Press, 2001.

Rantala, Kati and Sulkunen, Pekka: The Communitarian Preventive Paradox. Preventing Substance Abuse without the Substance. Critical Social Policy, forthcoming.

Evaluation in welfare is about changes in people's lives

Göran Sandell

Evaluation, evidence, quality and user involvement are becoming increasingly common concepts in everyday welfare practice. The need for evaluation, evidence-based information and quality assurance has been put on the political agenda, enshrined in legislation and recommendations, and addressed in national projects. In addition, the media has begun to take interest in the issue, and users are increasingly voicing their concerns. However, the increase in popularity has also added to conceptual confusion around the objectives, knowledge interests and areas of application of evaluation. It is not clear whose interpretations are to be given precedence and who has the right of evaluation, nor is it clear how evaluation is related to quality assurance, service development, user involvement, internal and external evaluation and information gathering. There are also ideological differences as to whether the focus should be on performance and economy or on real changes in people's lives. The issue is complex and difficult to grasp, but also enthralling.

I would therefore like to outline my views of evaluation in welfare, i.e. to discuss what can be evaluated, what is being done and what is missing, who can evaluate what and for what purposes, what the knowledge interests involved are and what genuine user involvement means.

What evaluation means in the context of welfare activities?

Generally speaking, evaluation means the process of assessing the significance, outcomes or benefits of different decisions, methods, interventions or activities by examining and clarifying them in terms of changes, distributions and priorities. In this basic sense, evaluation can, in principle, be defined as the examination and assessment of the outcomes and impacts of activities, decisions and interventions, regardless of the area of activity concerned.

Welfare activities are typically somehow targeted to those people who, depending on the type of activity, are referred to as users, clients, patients, care recipients, etc., or in a broader sense, even on their families and social networks. In the context of welfare activities, evaluation can thus be generally defined as referring to the examination and assessment of the outcomes and impacts of activities, decisions and interventions in the lives of the people concerned.

Changes, such as the outcomes of interventions, can be examined afterwards, i.e. after the intervention, but in order for us to be able to see what kinds of changes have taken place, we also need to know what the situation was before the intervention. Therefore, to identify the changes that have taken place during the evaluation period, it is necessary to analyse the life situation of the people affected by the intervention not only after but also before the intervention. The underlying reason for the implementation of welfare interventions is that there are needs for care, treatment and change that people confront in one or more areas of life in their life situations, and

these needs cannot be entirely satisfied with the individual's, family's or network's own resources. To meet these needs and to bring about positive changes, people's own efforts need to be complemented with welfare interventions. The extent to which such interventions help people meet the need for change, thereby enhancing their own resources and improving their life situation, can be defined as the benefit of the intervention.

Furthermore, it is necessary to analyse how outcomes, actual changes, needs fulfilment and benefits relate to the overall objective of an activity. The objective of welfare activities is linked to the original reason for and the ultimate purpose of the pursuit of these activities; in most cases this objective is for the interventions to bring benefit and yield outcomes that meet people's needs for change. To avoid mixing up such needs for change with 'service expectations', they can be defined as goals that concern desired changes in the lives of the people concerned. Evaluation can thus also seek to examine, based on goals so defined, whether the outcomes of interventions have actually benefited the targeted people in the intended way. For instance, the objective of an intervention for treating drug abusers can be defined as an aspiration to help clients develop their personal resources so as to be able to live more independently and in a more responsible way, with less substance abuse, improved family relationships and meaningful employment. The task of evaluation could then be to examine whether the interventions that have been carried out have actually been of benefit in the sense that the defined goals for change have been reached in people's lives, and to explore to what extent this benefit accords with the overarching objective of the activities.

People's life situations, goals for desired change, actual changes, goal attainment and benefits can be studied via a wide variety of methods. In principle, however, it is always necessary to have some kind of dialogue between the user and professional staff in a common process that seeks information, since outcomes relate to changes in users' lives. When it comes to evaluation in welfare, it is also important to specify the overall welfare objectives against which the evaluation findings are to be assessed, as there may be several conflicting objectives. These objectives can be related to care, the search for knowledge, integration and emancipation, but they can also be related to social adaptation and repression. Furthermore, evaluation in welfare requires that the information gained should be such that it can be reintroduced into practice for continuous quality improvement aimed at benefiting users in the long run.

Evaluation in welfare can thus be defined as the examination of the outcomes and benefits of interventions in people's lives, the assessment of the fit between the actual benefits and the objectives of the intervention, and the reintroduction of information gained for use in quality improvement to the benefit of the people concerned.

What can be evaluated in the welfare context, what is being done and what is missing?

In principle, there are five aspects of welfare activities that can be evaluated:

Implementation of decisions on the internal resources of services, i.e. personnel, funding, facilities, administration and different kinds of operational performance

There are a great number of evaluations that focus on the resources and performance of services, thus only providing information on what is being done internally within the service. For instance, a service may include decision-making on the use of a new form of care in drug treatment. This decision can be evaluated e.g. by examining whether the necessary facilities and personnel have been made available and whether the work is going according to plan with some drug addicts, and by assessing performance by giving an account of reports, meetings, bed places, treatments, etc., during a specific period of time, as well as related costs. Evaluation of the internal resources and performance of a service provides information on a specific project. On the other hand, no information is gained on the clients affected, such as whether they have reduced their drug abuse, whether these changes correspond to the goals set and whether there is any association between these changes and the interventions. This information is very much of the same type as that provided in ordinary activity reports, and it can be questioned whether such studies can at all be called 'evaluative', as information is provided on neither the clients' lives nor the relationship between interventions and outcomes.

User experiences of service encounters, interaction and service quality.

There are plenty of evaluative studies that focus on users' experiences of service encounters and the way they have been treated, using both less profound client satisfaction surveys and more in-depth individual and group interview studies in the areas of both social sciences and nursing sciences, as well as various creative methods in quality improvement projects. Although information on user experiences of service encounters and staff attitudes is of importance, it primarily focuses on user perceptions instead of the outcomes of a particular intervention.

Implementation of various professional interventions and methods regarding analysis, support, care and treatment aimed at individuals, families and target groups

If information is to be obtained on the outcomes of certain interventions, such as a specific method for family work, it may appear relatively simple to acquire it e.g. by asking the family about any changes and about their perceptions of service encounters and the treatment they have received. Such results, however, can be misleading. As there is no clear causal relationship between intervention and outcomes, and as human life is complex and multifaceted, with one and the same family possibly being subject to several interventions at the same time, it is difficult to find out whether it is one particular method that has led to the changes. This type of outcomes evaluation is carried out to a limited degree mainly in the context of evidence-based approaches advocating randomised controlled trials.

Implementation of regular welfare activities, including all interventions

Increasing information about people's life situations and their needs for change and improving understanding of the relationship between interventions and outcomes within the diversity of regular welfare activities, such as social welfare services within a municipality or rehabilitation services within a province, can involve analysing a number of complex and comprehensive projects. To achieve this, information should be gathered on all users and on all interventions and outcomes, and then analysed to reveal any explanatory interrelationships. However, hardly any evaluative studies in welfare have been conducted in the Nordic countries or elsewhere to find covariation and associations between interventions and outcomes in the whole area of regular welfare activities.

Social and health policy decisions and inputs in different welfare sectors with reflections on social development and the 'public good'

It is also possible to evaluate activities by examining the effects of organisational changes or external social and health policy decisions. If the analysis is confined to the different internal aspects of the organisation itself before and after the organisational changes, the information will, of course, cover these aspects only. If we also want to know how organisational changes or social policy decisions have affected the possibility of achieving operational objectives, it is necessary to explore the life situation, needs and resources of the people involved both before and after these changes and decisions, as well as to analyse covariation between these and potential changes in people's lives. Such areas of study as pedagogy, politics and economics have a rich tradition in evaluative research, which has in most cases focused on various organisational, political and economic aspects, while the implications and benefits for users have more seldom been the focus of interest.

Furthermore, it is possible to examine the effects of events taking place 'outside' the actual activity. For instance, the effects of economic fluctuations on the financial situation of different groups can be studied, and it can be further investigated how this influences the provision of financial assistance by social welfare authorities. However, such evaluations are not carried out with regard to any explicit welfare objective, and therefore this type of knowledge, though important as such, can hardly be considered evaluative.

To sum up, there are thus several studies of the internal aspects and performance of activities, of the user's experience of service encounters, and of various organisational changes and social and health policy influences, in which the focus of evaluation is on activity, policy and economy. However, there is less evaluative research which aims to link interventions with outcomes and benefits, although there is some serious research in the tradition of both logico-empirical evidence-based knowledge generation and hermeneutic social sciences. On the other hand, there are hardly any evaluative studies that incorporate all of those diversified regular activities that are going on in different welfare sectors in municipalities and provinces that would focus on affected users' life situation, needs for change and actual changes, and aim to link the effects of particular interventions with goal at-

tainment, user benefits and welfare objectives and to reintroduce the gained information into practice for continuous quality improvement.

Who can evaluate what in the context of welfare activities?

It is a widespread but simplistic view that for an evaluative study to be regarded as scientifically serious and credible, it needs to be carried out by an external researcher who investigates the activity from the 'outside', and that such studies should in no circumstances be carried out by someone who is employed in the service to be evaluated. This belief is shared by most researchers, but it is also common among welfare staff. With some exaggeration, the underlying justification could be presented as follows:

– "People who work within a particular welfare service are certainly competent and committed, but this is the very reason why they are likely to be blind to both the inner operations of the service and to client contacts. As they obviously want to act in the best interests of their clients, they will cause, consciously or not, a systematic bias in the material so that the outcomes will look better than what the reality is. And even if they have professional competence, they mostly have no knowledge of research methodology. They are likely to be led astray by what they regard as outcomes, and will thus draw false conclusions that cannot be generalised. Moreover, there is no reason to expect that they would be willing to report any negative outcomes, as who would chop off the branch upon which one actually sits. And if the evaluation should reveal shortcomings in the management of the service, the staff would hardly have courage enough to report these shortcomings in an era of restructuring and individual pay determination.

– On the contrary, evaluation should be carried out by a researcher with a good knowledge of methodology, independent of both those in positions of power and the clients, and capable of openly reporting both positive and negative outcomes without any risk of becoming personally involved. It might be possible for staff to gather information, for instance, based on a tested standard questionnaire, but independent researchers should in any case be responsible for the design, method, analysis and conclusions. Nothing else will guarantee reliability."

This view is based on ignorance and misapprehension. People do not seem to be aware that different matters can be evaluated with different objectives as a starting point and from different perspectives and with different knowledge interests, which in turn is bound to influence the position and relationship those in charge of data collection will assume with regard to the 'object of study'.

If the knowledge interest is oriented towards control, i.e. the perspective is that of policy-makers, management or supervision authorities, it is self-evident that staff cannot act as information gatherers. This is because they are in a subordinated position dependent on the management and thus cannot control or supervise 'themselves'.

If the knowledge interest is to call into question and disclose, i.e. the perspective is that of science or the media, where the focus of interest is, for instance, on the impact of the exercise of power on staff's psychosocial work environment, it is not advisable that people in a managerial position or their

subordinates carry out the data collection. Neither is it appropriate that staff are assigned the task of investigating how their own clients experience encounters with the service and staff attitudes in the care relationship. As care providers, the staff constitute one party of the relationship to be studied, which means that, in principle, there is always an in-built unequal power relation involved. For the same reason, it is not recommended that staff should ask their own clients if they are 'satisfied'. – The situation is the same as if I asked someone who is dependent on me "Am I a nice person?", which would in all likelihood produce a very biased result.

Accordingly, it is inappropriate for staff to act as information gatherers if the knowledge interest is to control or to disclose, with an 'inward' focus on the resources of the service concerned, i.e. if the intention is to study the inner operations of the service and encounters with it.

However, the logic is reversed if the knowledge interest is informative, whereby the intention is to explore the affected people's life situation and needs for change, the interventions to be carried out and the changes that actually have taken place in these people's lives. In this case, the evaluation is focused 'outwards' with the aim of relating interventions and outcomes to the objectives and benefits of the service from the users' perspective. Here there can be reason for suspicion about external evaluators who usually lack the professional competence, experience and commitment that the staff have. General practitioners, school health nurses, child psychologists, social work secretaries, practical mental nurses, home service personnel and others are certainly more knowledgeable about their patients, clients and other care recipients than a visiting researcher or temporarily employed research staff. It is true that welfare professionals in most cases do not use any documentation structure (which would allow group-level collection of information with more generalisable conclusions) – but it is usually possible for committed staff to gain plenty of knowledge about the people concerned in their 'own' individual cases. Information collection in the course of regular work also enables different matters to be studied repeatedly on different occasions, and misunderstandings to be cleared up.

Furthermore, it is hardly likely that professional staff are led astray and unknowingly make people's life situations look better than what they are. It is simply presumptuous and a false myth to claim that staff present or fabricate too positive outcomes due to their excessive commitment to their clients or because they are afraid of negative consequences for their own work situation. Many years of experience in developing the integrated model show this not to be the case. On the contrary, experienced welfare professionals have usually adopted a critical and suspicious attitude towards any outcomes that are achieved too easily or too fast. This is because they have seen that there are sometimes no changes at all, and because the outcomes are not always permanent. Also, it is an element of their survival strategy to avoid repeated disappointments with regard to their professional commitment. Therefore, the risk is not that staff will report unrealistically good outcomes, but rather that they are to a degree 'cautiously negative', just to be on the safe side. In addition, staff are well aware that it would look strange, for instance, to report a number of 'successfully completed cases' at the end of year – and early next year record them again as 'still outstanding cases'. Of course, nothing is gained in long-range welfare work by attempts to make

the outcomes look better temporarily. It is in fact more likely that a temporary, supposedly neutral and objective visitor to a specific field of activity (regardless of the visitor's level of education) has more difficulty with regard to both the trustworthiness of the users studied and the truthfulness of factual information. Accordingly, it is both feasible and desirable that professional staff gather the information if the knowledge interest is informative with an 'outward' focus, i.e. the intention is to examine the effects and benefits of activities for people's life situations.

To sum up, it depends on the objectives, focus and knowledge interest of evaluation as to who can evaluate what in the context of welfare activities:

- If the evaluation is carried out to control or disclose internal aspects of a service, information should be gathered by people external to and independent of the service.
- If the evaluation is carried out to improve the service by gaining a deeper understanding of what is happening in those people's lives who are affected by the service, the required information should be gathered by competent professional staff who preferably are employed within the service.

What is meant by genuine user involvement?

Evaluative studies can be roughly divided into 'inward' evaluations focusing on the organisation's internal activity, funding and performance, and realistic evaluations focusing on the effects of activities on people's lives, for instance, through outcomes analyses for evidence-based information. Both of these have developed from more distanced and objectivising approaches towards an increasing interest in 'processes', regarding both what is happening within the activity, how different services are 'delivered' to the people concerned in their roles as 'clients' and 'users', and how these people as service recipients are treated in their encounters with the welfare system. Ideologically differing evaluation traditions have gradually developed.

Users as clients and citizens and the risk of negative paternalistic attitudes being replaced with a simplistic user perspective

One ideological approach is to underline the rights and freedom of choice of the individual and to see welfare activities primarily as service management organisations that deliver a range of services at certain prices on a more or less competitive market to more or less satisfied 'clients'. Doubtful attitudes are mostly expressed towards an excessively 'strong society' which tends to foster paternalistic attitudes towards the individual, thus giving rise to 'learned helplessness'.

Instead, it is argued, official control should be relaxed in the context of welfare and the individual client should be given more control over benefits, such as care and education vouchers. Through their own choice and purchase of welfare services, individual clients can exercise influence as consumers as they see fit. Evaluation thus comes closer to consumer empowerment with its studies of client satisfaction, mostly through client surveys. Evaluation seeks to find out what the client wants, how expectations are met, whether the price is competitive and how clients perceive services and the way they are treated.

The Nordic welfare legislation has knowingly distanced itself from earlier authoritarian language and paternalistic attitudes. Existing health, child, family, disability, elderly care and psychiatry services all underline the importance of strengthening individual citizens' rights, enhancing personal resources and taking account of user views and experiences in encounters with the welfare system. In practice, however, there is the risk that welfare organisations still foster paternalistic attitudes when they make judgements of their clients behind their backs and over their heads, when they conduct studies where the emphasis is on users' problems and inadequate capabilities, and when staff and policy-makers take socially adapted attitudes towards normalised behaviour.

Another ideological approach is to advocate, for the above reasons, a more democratic perspective, with an emphasis on justice and solidarity, and to stress users' empowerment in their citizen role. In this approach, welfare organisations are assessed based on the values enshrined in welfare legislation, such as equality. Efforts are made here to improve one's knowledge through a variety of individual and group interviews where even users can play an active role in the formulation of criteria for what equality means in this context and to what extent they have received equal treatment in their encounters with the welfare system. An interesting model for this type of user involvement, developed by Hanne Krogstrup, is presented in this publication.

However, there is a risk involved in the use of diverse user opinions on the welfare system as a basis for evaluation. By way of definition, users' personal experiences provide biased information and tell us nothing about the internal and external changes that actually have taken place in these people's lives. Information can only be gained about users' experiences and views of service delivery, and not on the actual changes in the life situation of the user group as a whole. This kind of model can thus be extremely useful for process-oriented user studies fostering democracy and aimed at quality improvement. However, they are not useful in activity evaluation, as no information is obtained on life situations, actual changes or the relationship between interventions and outcomes.

The fact that such user involvement has certain limitations does not, of course, prevent users from exercising influence more generally, both through participation in various interest organisations and in their citizen role through the democratic electoral system.

In principle, there are thus at least three different kinds of doubts expressed about user involvement:

– users' needs are reduced to customer expectations on a supposed welfare market;
– assessments are made by staff from a von-oben perspective over the users' heads;
– a positive but unconsidered aspiration for democracy leads to an unreasonable emphasis on a few sporadic users' opinions instead of staff's professional knowledge and experiences.

Users are human beings with needs for change and personal resources.

Welfare professionals do understand that users in certain respects can be regarded as customers who may reject a specific welfare offer; in general, the underlying welfare ideology is defended; and in most cases users' perceptions of interventions are constantly appraised together with the users themselves. However, the fundamental reason why welfare staff establish contacts with users is that these are human beings with various needs for care and needs for change and that the welfare system's overall objective is to provide help and meet these needs in consultation with the users. Accordingly, users cannot simply be regarded as members of the public, citizens of an electorate or customers who select an appropriate level of services at a certain price.

The slogan that 'the customer is always right' (in the naive service management sense) does not always hold true in users' encounters with the welfare system, as users are seldom aware of or able to express their 'real' needs. Simple service interventions, such as a window-washing service for a certain price in the context of general home-help assistance, can possibly be regarded as pure customer-supplier relationships. As soon as core welfare functions are involved, i.e. the intention is to help people meet their more fundamental needs, users nearly always need to receive help from competent, committed staff in defining their needs and finding appropriate treatment and care measures and interventions for change. This is one of the basic reasons why there is a need for professionally qualified physicians, psychologists, social workers, nurses, care givers, etc. in the welfare sector, instead of people who only read answers from questionnaires or act as 'sellers' of welfare services.

The primary reason for users' encounters with the welfare system is certainly that they want to develop their resources and satisfy their needs so as to be able to live a healthier life with a higher degree of need satisfaction. In the context of welfare provision, the user perspective cannot therefore be confined to 'the public', 'voters', 'customers' or 'service recipients'. Instead, it has a deeper meaning, referring to human beings with needs for change and personal resources for change.

The way users are treated, their needs investigated and interventions for change carried out are of decisive importance for ethical considerations, co-operation and outcomes. A respectful, emphatic and salutogenic attitude generally contributes to gaining good outcomes. This is why users' experiences and perceptions of service encounters with and interventions by welfare organisations are also important, although it is reasonable to use actual changes in people's lives as the overall evaluation criteria. Welfare practice, though, is not a matter of choosing between good service encounters or good outcomes. On a deeper level, views of human nature, attitudes, interventions, competence, co-operation and outcomes are always mutually dependent.

Genuine user involvement through co-operation and dialogue with people

User involvement in evaluation is concerned with efforts to build up real co-operation with users in a profound and radical way, in searching for infor-

mation about needs, resources and changes and the ways in which this information can be put to use in jointly agreed interventions for change. The prerequisite for such genuine user involvement is thus that it should become an integral part of the direct regular work carried out in established welfare organisations to bring about change, and that evaluation in turn should be integrated in this work for change, in constant dialogue between users and staff. User involvement cannot thus be restricted to a specific project limited in time and scope and concerning a limited number of articulate users, but it should become an integral part of all regular ongoing welfare activities and involve all users. The information that can be gained through this kind of integrated dialogue can provide broader views and deeper insights into what is possible through separate assessments by users and staff.

The main reasons for user involvement in evaluation are thus that in addition to gaining more in-depth information about people's needs for change and being more likely to be able to make use of this information in individual and target-group-oriented work to bring about change together with the users themselves, user involvement contributes to enhancing users' personal resources through a salutogenic attitude where knowledge is sought in a common dialogue.

User involvement can thus increase knowledge in a number of ways: on people's life situations and needs for change, on interventions for change and co-operative relationships with the people concerned, on outcomes and goal attainment in their lives, and on long-term user-oriented service-quality improvements. Furthermore, when ongoing welfare activities are evaluated in genuine interaction with users, it is not possible to analyse and deal with a specific case whilst taking a particular attitude and to then evaluate it with some other attitude in another situation. In the integrated evaluation approach, all these different aspects can be seen as different stages in one and the same welfare activity, which means that the traditional division into analysis, diagnosis, treatment and evaluation has been abandoned on the mental level. Integrated evaluation thus becomes a part of regular welfare provision, as well as analysis and treatment.

Dialogic interviews can be integrated into regular work routines and carried out by professional, committed, culture-competent welfare workers, thereby allowing a long-term, trustworthy personal contact to develop, with a broad knowledge of background and context that will naturally increase the validity and reliability of the findings. In addition, the dialogue form promotes empathy and confidence, and provides both users and staff with an opportunity for metacommunication in and about the interview situation itself (i.e. to be able to discuss what is being discussed, what the participants really mean, feel etc.) to increase mutual understanding.

Dialogue thus allows an understanding to be gained of people's life situations and needs that is both more reliable and deeper than what can be obtained by means of observations or surveys and interviews using structured questionnaires, regardless of whether these are carried out by staff of more 'unbiased' external interviewers. Moreover, in the context of welfare work, validity, which is concerned with the degree to which the study focuses on what is relevant and necessary for adequate understanding of what is to be investigated, can be controlled and confirmed by colleagues through co-operation and supervision, by client records, etc.

There is one exception to this principle – the client's view of the relationship between himself or herself and the staff member conducting the interview. The risk is that the structurally unbalanced power relationship in the encounter between the client and the social worker/the patient and the doctor, etc., will limit the reliability of the answers. (This relationship can be more properly analysed using such methods as colleague reviews, focus groups and external interviews).

Genuine user involvement is thus a matter of common efforts to bring about change, and evaluation is an integral part of this work, in which users themselves can actively participate. Reliable and valid information about people's life situations, needs for change, interventions, changes and benefits can be gained in a structured dialogue and integrated via an ongoing and continuous co-operation between professional staff and users.

References

Bateson, G. (1987). Ande och natur, en nödvändig enhet. Lund: Symposion.

Cheetham, J., Fuller, R., McIvor, G.& Petch, A. (1992): Evaluating Social Work Effectiveness. Buckingham: Open University Press.

Clarke, A. (1999). Evaluation research - principles, methods and practice. Thousand Oaks, Calif. ; London : Sage.

Fuller, R. & Petch, A. (1995). Practitioners Research. The Reflexive Social Worker. Buckingham: Open University Press.

Gomm, R. & Davies, C. (Eds.), (2000). Using Evidence in Health and Social Care. London: Sage.

House, E. R.(1999). Values in evaluation and social research. London : Sage.

Karlsson, O. (1999): Utvärdering - mer än metod : tankar och synsätt i utvärderingsforskning. En översikt. Stockholm: Ajour, Kommentus

Kazi, M. (1998). Single-case evaluation by social workers . Aldershot : Ashgate.

Kazi, M.A.F, Blom, B, Moren, S, Perdal, A.-L. & Rosila, L. (2002). Realist evaluation for practice in Sweden, Finland and England. Tempere: Artikel utvärderingskonferens.

Owen, J. M. (1999). Program evaluation. Forms and approaches. London: Sage

Patton, M.Q.(2002).Qualitative research and evaluation methods.ThousandOaks: Sage.

Pawson, R. & Tilley, N. (1997). Realistic evaluation. Thousand Oaks: Sage

Reamer, F. G. (1998). Social work research and evaluation skills : a case-based, user-friendly approach. New York, Columbia University Press.

Robson, C. (2000). Small-Scale Evaluation. London: Sage.

Sandell, G. (1985). Psykosocialt förändringsarbete. Doktorsavhandling, Institutionen för socialt arbete, Göteborgs universitet samt bok på Liber förlag, Stockholm.

Sandell, G.(1998 – 2003). Integrerad utvärdering. Ideologi, teori, metod och IT-stöd. Paradigm och modell för utvärdering och kvalitetsutveckling inom välfärdsverksam-heter. Bok och bildspel samt Info till dataprogram / Analysity.

Scriven, M. (1991) Evaluation thesaurus . Newbury Park, Calif.: Sage.

Shaw, I. & Lishman, P.(1999) Evaluation and Social Work Practice. London: Sage

Vedung, E. (1998). Utvärdering i politik och förvaltning. Lund: Studentlitteratur.

Evaluation From a User Perspective

Hanne Kathrine Krogstrup

The editor originally asked me for a contribution entitled 'User Participation in Evaluation'. It should, however, be noted that in the following article the concept of 'user' designates users of public social services, while other authors using the concept refer to users of the evaluation.

Over the last decade, user participation has been placed on the agenda in many contexts and also in relation to evaluation. The reasons for user participation in evaluation are based on several overlapping arguments. In this context, four arguments for user participation are discussed: a control argument, a democratic argument, a knowledge argument, and an emancipatory argument.

It should be emphasized that these forms of participation are not ideal-typical. In other words, the typologies are not mutually excluding. Embedded in the individual evaluation model, we may very well find, for instance, both a control and a knowledge argument. In the following, each of the four arguments for involving users is analyzed in detail.

Before turning to the analysis, I wish to present a different analytical distinction in the understanding of user participation in evaluation. That is, the distinction between top-down and bottom-up evaluation (Krogstrup 2001). The control argument is linked to a top-down understanding and possibly to an assumption about best practice. The democratic argument, the knowledge argument, and the emancipatory argument are linked to a bottom-up understanding. The distinction between top-down versus bottom-up is determined by which link in the parliamentary chain determines the criteria for evaluation.

Assigning the users a specific role in the described evaluation models is not a value orientation which points toward the only object of the social sector being to consider the needs and particular interests of the users. The objectives of the social sector, both societal and institutional, are to safeguard the common good and particular interests. At the institutional level, the common good must be considered as well as the particular interests of the users to the full extent possible within the institutional frameworks which are subject to financial, legal and legitimacy constraints (Otter 1985). The considerations express political prioritization between the common good and the interests of particular groups to which the social security legislation extend.

The Control Argument

This argument is most clearly reflected in the new Public Management wave. Here the users are involved through, for example, the joint determination by the administrative management and employees of the criteria for evaluation, and, subsequently, measuring to what extent the criteria have been met. One example is the so-called user satisfaction surveys. The major

task of the users is to fill in a questionnaire, the purpose of which is to assess whether the criteria have been met. This type of user participation is often referred to as top-down oriented. In other words, it is the top level of the organization that determines the criteria (Krogstrup 2001).

Surveys of user satisfaction occupy a strong position especially in organizations that have systematized evaluation as their common recipe for the form and frequency of evaluation. Such surveys often use a range of indicators for measuring user satisfaction. Furthermore, these surveys are, in many cases, embedded in the organizational routines of large organizations operating with evaluation units and specialists employed to cater for that activity (Dahler-Larsen 2002:47).

A typified process of measuring user satisfaction can be described as[10]:

> The political level determines the economic framework and the general objective of the endeavour (e.g. the administration must have a 'personal face' vis-à-vis the citizen). Note that the precondition of measurement is clear objectives.
>
> The management, or other determinators, then define performance criteria and standards for realizing the objective. The criteria are operational, concrete measures of quality; such as disabled persons must receive service from a specific social worker in the administration (e.g. 70% of the users must have received service from a 'personal social worker').
>
> Performance criteria (the concrete objectives of quality) are translated into a questionnaire directed at the users or measuring quantitatively to what extent they perceive the services satisfactorily (e.g. that 60% of the users meet with the same social worker over a three-month period).
>
> Against this backdrop, the administration evaluates whether a possible gap in quality exists between organizational standards and user assessments (that is a gap between the standard of 70% and the quality assessment of 60%).

In effect the organization can attempt to bridge the quality gap (KL, Sundhedsstyrelsen, Leth, Nørgaard Madsen, Grönroos in Krogstrup 1997b:68)

User satisfaction surveys may be modeled differently from this description - that is, the users may have contributed to generate the criteria through prior qualitative interviews. The formalized description of user satisfaction surveys are indicative of a monitoring system, and in this context the following possibilities and constraints are traditionally emphasized.

[10] It should be noted that the survey demonstrating user satisfaction surveys to be of significant motivation does not enable detailed specification of the concept and is therefore not necessarily in keeping with the ideal-typical description.

Possibilities

The immediate advantages of user surveys are emphasized as the inclusion of the users and that employees in the public sector are held responsible for meeting the preferences of users. User satisfaction surveys are also viewed as beneficial in relation to assessing to what extent the social work complies with the criteria determined politically or managerially. That is, are the users receiving the service that they have been promised politically? Criteria for user satisfaction surveys are often based on so-called quality standards that in a Danish context are communicated by way of service declarations. Hence, user satisfaction surveys are applicable in controlling responsibility in tasks performed, and possibly constitute the basis for an in-depth assessment of the survey results.

Constraints

The critics point toward user satisfaction surveys as not providing a valid image of user satisfaction. The argument is that the users are responding to questions that those determining the criteria consider relevant. But these questions are not necessarily in accordance with those the users conceive as relevant. Persons (managers, officials, professionals, etc.) without real knowledge of social work (often because the political objectives are vague or not operational) are often determining the criteria. In effect, those that determine criteria will indirectly become policy formulators (and thereby indicate the direction for the 'good" social work) without being held responsible (for a detailed criticism see Krogstrup 2001b; Krogstrup & Tjalve 1999).

Social work often involves such complex processes that user satisfaction surveys appear much too simplified. The argument is increasingly being advocated that user satisfaction surveys must be founded in evidence-based knowledge and best practice: Best practice has become the most sought after form and knowledge. Not just effective, decent, or better practice, but best practice. It is the American way. Be the best you can be. How? Learn lessons (local knowledge about what works) and convert them into best practice (universal knowledge about what works, at least by implications and being best) (Patton 2001:330). Patton summarizes the criticism of best practice and hence also the criticism of user satisfaction surveys as: 'Seldom do such statements identify for whom the practice is best, under what conditions [it] is best, or what values or assumption undergird the best-ness' (ibid.:330)

The Democratic Argument

In the context of evaluation, Ernest House and Kenneth Howe have presented the strongest and clearest democratic argument in their introduction of deliberative democratic evaluation. The democratizing role of evaluations

was a relatively salient theme at three international conferences in 2000[11]. Here it was advocated that the top-down oriented control logic of the western societies contributes to the creation of a democratic deficit which must be compensated for by the inclusion of the users in dialogue-oriented models.

Democratic deliberative evaluation is not a genuine evaluation model, but rather a normative statement of intent about what good evaluation practice should include. The American scholars Ernest House and Kenneth Howe have conceptualized these statements inspired by 'Nordic thinking' within evaluation, and they are explicitly referring to the Swedish professor in evaluation, Ove Karlsson. Ernest House is very critical toward evidence-based knowledge which he finds to be advocated much too strongly – not as a science but as an ideology in the realm of the New Public Management strategy. House stresses that a one-sided evaluation strategy aiming at generating evidence-based knowledge excludes other forms of knowledge, such as the knowledge that democratic deliberative evaluation might generate, and leads to methodological fundamentalism in evaluation (House 2003).

Deliberative democratic evaluation was introduced in 1999. House and Howe stress the following ten questions as being important for determining whether a deliberative democratic evaluation is exhaustive: 1) What are the interests represented? 2) Are important stakeholders represented? 3) Are some excluded? 4) Is the imbalance of power serious? Have processes been established for controlling imbalance among stakeholders? 6) How are people participating in the evaluation? 7) How authentic is their participation? 8) How complex is their participation? 9) Is reflexive deliberation taking place? 10) How reflected and comprehensive is the deliberation? (House & Howe 1999).

Three requirements are defined for the evaluation design: Deliberation, dialogue, and inclusion. Deliberation is defined as reflexive reasoning about relevant themes, problems, and questions. The purpose is to identify the preferences and values of the stakeholders. The approach is dialogic in the sense that stakeholders and evaluators are engaged in a dialogue throughout the entire evaluation process. The purpose is to portray the attitudes and ideas of the stakeholders as fully as possible. In relation to a discussion about user inclusion in evaluation, the requirement of inclusion in particular becomes interesting. According to House and Howe (2000) inclusion requires that 'Evaluators must design evaluation so that relevant interests are represented and so that there is some balance of power among these interests, which often means representing the interests of those who might be excluded from the discussion because their interests are most likely to be overlooked. And of course deliberation should be based on the discussion of merits, not on the social status of participants' (House & Howe 2000:6).

House and Howe see deliberative democracy as an ideal worth pursuing, and even though it is difficult, 'the lack of perfection is no reason to stop trying to do the best we can' (ibid.:9).

[11] Canadian Evaluation Society, Montreal May 2000 (Theme: Evaluation and the New Governance). The 2000 EES Conference, Lausanne, October (Theme: Taking Evaluation to the people: Between Civil Society, Public Management and the Polity). American Evaluation Society, Hawaii, November 2000 (Theme: Increasing Evaluation Capacity).

Possibilities

The approach aspires to attain democratic participation for all stakeholders at all levels. This is seen as positive in relation to the value-based argument that the prime task of evaluation is to secure the equal participation of all stakeholders. The core of this democratic perspective is for the evaluator to be open toward the views of all stakeholders and assume responsibility for conducting the assessment of the program in a justifiable way. That is, it is left to the evaluator to generate the questions for the evaluation as opposed to what happens in the stakeholder model, where questions are generated on the basis of consensus among the evaluator and the involved stakeholders (Stufflebeam 2001:74-76). Like the other responsive approaches, stakeholder involvement is expected to further the utility of evaluation results.

Constraints

House and Howe point out, however, that their evaluation is unrealistic and hardly fully applicable in relation to the ideals. The perspective is based on full democratic participation. As an example, Stufflebeam emphasizes that organizational changes according to the premises of the evaluation would require consensus (Stufflebeam 2001). House and Howe (2000) respond to this criticism by saying that they are aware of the fact that their approach is not a model but an ideal that should be pursued.

A different criticism of the approach is that it allows the evaluator too much discretion in excluding input assumed to be unethical or incorrect. In other words, the evaluator's interpretation is decisive in determining which questions are selected for evaluation.

Furthermore, a number of scholars are critical of the democratic argument. Stake (2000:97), for example, mentions that dialogic, deliberative, and dialectic processes in general are under-prioritized in practical evaluation work (in an American context), while simultaneously arguing for the primary task of program evaluation as being to assess the advantages and imperfections of services and not to promote democracy. The task of the evaluator is thus to generate information that facilitates the understanding of programs, accountability and to contribute to political decision-making processes. Furthermore, evaluators should support ethical, professional, cultural values, etc. Stake thus refutes the ideal of the evaluation to promote democracy, but he concedes that democracy is sometimes a secondary purpose of the evaluation (Stake 2000).

The Knowledge Argument

The knowledge argument primarily refers to the apparently growing problems of the Danish welfare state in resolving social tasks. The alleged reason is that the public sector 'solutions' match inadequately the social problems. The purpose of involving the users is, in this context, to allow for the rationality of the users in organizing public services to ensure accuracy or

goal realization (Rothstein 1994), in particular in relation to dynamic interventionist solutions.

The UPQA model

The UPQA model (User Participation in Quality Assessment) described below corresponds with the knowledge model. The model contains four major stages:

Stage 1:
Group interviews in which the users are asked to express and justify which aspects of the services offered are positive and negative, respectively. The result of this group interview will be a series of experiences with and subjective assessments of quality.

Stage 2:
Front stage employees are presented with the users' opinions to be discussed in a group interview in order to get their opinion on the reasons for the users' evaluation. The major purpose is to make the front stage employees reflect upon their practice.

Stage 3:
Statements from the group interviews with users and front stage employees are now presented to the management of public service to be discussed in a group interview in order to get its opinion on the reasons for the users' and the employees' statements.

Stage 4:
Statements from the group interviews with users, front stage employees, and the management of civil service administration are now presented at the political level to get its assessment of the reasons for the statements of these actors.

The interview with the users triggers an evaluation and learning process at the other levels of the municipal organization. The purpose is to challenge existing norms governing practice and action. In effect the users' understanding of the problems and the social work at different levels in the organization are made to agree much better (Krogstrup 1996, 1997a).

Possibilities

Like in the deliberative democratic model, the UPQA assigns a particular position to the users, in this context 'triggers for learning'. Embedded in this understanding is a value orientation determined a priori that must not be confused with the users as determining the evaluation criteria. Evaluation criteria are generated in the evaluation process though the inclusion of various stakeholders, holding their different roles and tasks in respect. The model strongly emphasizes the triggering of a learning process through con-

fronting the stakeholders' interpretations of and perspectives on a given service. The learning taking place will probably be most pronounced among the social workers.

The UPQA model respects the value pluralism of public organizations which should be reflected in the evaluation and its results, and it thus 'rejects' the idea of being able to generate unique evaluation results (best practice). To the extent that data generated via the UPQA model are used as the foundation for designing user satisfaction surveys, the approach is bottom-up and embedding the users' criteria of relevance in the user satisfaction survey. (See the article User Evaluation in Practice). Nevertheless, a user satisfaction survey in this perspective will, in part, hold critical problems similar to those mentioned under the criticism of user satisfaction surveys.

Constrains

Experience shows that implementing the UPQA model involves at least two obstacles. First, people not familiar with evaluation find it difficult to manage the relatively large quantities of data that the UPQA model is capable of generating. Second, The UPQA process is often terminated before the management level is confronted with data; in other words, the evaluation process is from time to time limited to the inclusion of users and front line employees. One explanation is, as stated by four Danish municipalities working with the model, that the tradition for bottom-up communication in the control chain is not particularly well developed. Compared with, for example, a user satisfaction survey within evaluation, the UPQA model produces data from several levels of the control chain which in a 'best practice' context are understood as diffuse data pointing in several directions. Oddly enough Ernest House (the father of democratic deliberative evaluation) interprets the UPQA model as a democratic deliberative evaluation (personal conversation).

The Emancipatory Argument

This argument refers to the prime purpose of involving the users as liberating the unexploited potentials of the users. Empowerment evaluation corresponds with this argument.

Empowerment evaluation

Empowerment and empowerment processes are familiar concepts within the social realm and are often traced back to Paulo Freire. The assumption underlying empowerment evaluation is that every human being possesses individual unique competencies, interests, and needs that deserve a chance to unfold. The idea is that human beings deserve equal opportunities to express their unique potentials, and that no superior mechanisms ought to exist that sort people into categories defining their needs and interests. One precondition for man being able to assume responsibility for his own life is that he

has the opportunity of formulating the premises for this life. Inclusion into such a process must necessarily, to be fair, start from the capacity of the individual human being to make decisions and clarify self-defined needs and interests. The underlying premise for implementing empowerment evaluation is thus an understanding of individuals as dynamic and capable of developing the ability to change the conditions for their own lives rather than being static creatures of diverging physical, psychic and mental handicaps (Mithaug 1996:239).

David Fetterman presents empowerment evaluation as a process of four stages (Fetterman 1996:18-20) The first stage involves clarifying the strengths and weaknesses of a given service viewed from a user perspective or its possibilities for and constraints on the users' self-realization. The purpose is really not to evaluate the current state-of-affairs, but rather to create a baseline against which future changes and movements can be evaluated. In the second stage of the process, qualitative objectives for development are determined in such a way that coherence is established between daily activities and realistic objectives. In the third stage, strategies for accomplishing the objectives are developed. In the fourth stage the participants decide the type of documentation to be generated in order to assess whether the original objectives have been accomplished.

Possibilities

In certain situations the evaluator may function as counsellor to those involved. That is, assisting people, who are not in control of their own lives, to become empowered through the evaluation process. If this is the purpose, the evaluator must ensure, through active engagement, that the participants determine the guidelines for the evaluation and for the solutions to their own problems (Fetterman 1996:9-16).

Rather than functioning as an expert or a counsellor in relation to the political decision-making process, the role of the empowerment evaluator is to collaborate with stakeholders in the organization and to support the empowerment oriented self-evaluation process. Being part of this collaboration, the evaluator learns about the participants' viewpoints, and their views of reality and the difficulties of life. The purpose of the process is not a final assessment as is the case of traditional evaluation, but to develop and qualify self-reflection through the process of liberating the potential resources of the users.

Constraints

Some scholars fear that the evaluator being involved in development processes may make it difficult to meet the underlying requirement of rendering visible the evaluation results. According to Fetterman there is, however, no reason to dismiss empowerment evaluation, as the processes are capable of producing even richer data (Fetterman 1996:6) or can help qualify a more 'traditional' evaluation.

There are only a few examples in the literature of criticism of empowerment evaluation, but there are, of course, certain reservations. It is unclear how empowerment evaluation stands in relation to other types of evaluation, and the criteria are blurred as to what way empowerment differs from collaboration and inclusion (cf. Patton 1997). In a Danish context, it has been questioned whether empowerment, and hence also empowerment evaluation initiated by employees in the public sector, will lead to the expected results. The argument is that employees can hardly initiate emancipatory processes that are themselves challenging (lecture by Jørgen Elm Larsen, Copenhagen University).

Empowerment as a strategy for social work is relatively widespread in the Nordic countries. In this work several elements from empowering evaluation are identifiable though they are representing a method for social work rather than actual evaluation.

Conclusion

In the article, I have attempted to point out the possibilities and constraints characterizing each of the models for user participation. However, in doing so, I have not been rigorous in the sense of defining these uniformly. Assessing the possibilities and constraints of evaluation models diverges depending on the perspective and criteria applied to the evaluation model, meaning that this type of evaluation is not different from other types of evaluation. Nor are the underlined possibilities and constraints exhaustive in the sense that they are the only ones. But the possibilities and constraints suggested here are the ones most often referred to in the literature on evaluation models that include the users.

Many western countries, including the Nordic countries, have imported New Public Management concepts, though to varying degrees. This has had implications for changes in the scenery of evaluation in general (a movement from evaluation based on a multitude of methods toward audit and performance measurements) and for the inclusion of users in evaluation specifically. User satisfaction surveys are becoming increasingly important, legitimized on the assumption about ‘best practice' and evidence-based knowledge. The tendency is that demands for quality standards to form the basis of monitoring are increasingly being integrated as legislative demands. Additionally, the allocation of financial resources for implementing evaluations is increasingly conditioned by mainstream concepts being applied (e.g. Cochrane and Cambell). However, some scholars have certain reservations about these tendencies. This is, among other things, reflected in certain scholars and practitioners maintaining the relevance of being responsive to the users by way of dialogic forums, such as the intentions underlying democratic deliberative evaluation, the UPQA model, and empowerment evaluation.

Diverse ‘world views’ lead to specific constructions of knowledge and truth and thus also to ‘truths’ about the best method for evaluation. Some will point toward user satisfaction surveys being ‘the best way of evaluation’, while others will point toward empowerment evaluation as ‘the best way of evaluation’ to mention but two extremes in the debate about inclu-

sion in evaluation. Thus when it comes to what is ‘true’ and ‘false ‘ in evaluation, different knowledge regimes hold different perceptions (Phillips 1999:22). A knowledge regime is a particular way of talking about and understanding the phenomenon of evaluation. Simultaneously, with discourses struggling to gain supremacy, the image of ‘true’ and ‘false’ evaluations are continuously subject to negotiation.

In the Nordic countries there is a tendency to attach growing importance to defining standards for services within the sectors of health, social services and education. Including the users in evaluations is thus an appropriate tool for generating standards reflected in the rationality of the users.

References

Dahler-Larsen, Peter (2002): Evaluering kortlagt; Gylling; Systime

Fetterman, David M. (1996): Empowerment Evaluation: An introduction to Theory and Practice I: Fetterman, David M., Shakeh J. Kaftarian & Abraham Wandersman (red.): Empowerment Evaluation. Knowledge and Tools for Self Assesment & Accountability. Thousand Oaks, Sage Publications pp 3-46

House, Ernest (1993): Bush's Neo-Fundamentalism and The New Politics and Evaluation. In Politics of Evaluation, Uppsala, August 28-29, 2003.

House, Ernest & Kenneth Howe (1999): Values in Evaluation and Social Research; Thousand Oaks, London, New Delhi; Sage Publications

House, Ernest & Kenneth Howe (2000): Deliberative Democratic Evaluation I: New Directions for Evaluation; Spring 2000; Number 85; p 3-12; Jossey Forlag

Jørgensen, Marianne Winther og Louise Phillips (1999): Diskurs Analayse som teori og metode; Samfundslittaratur, Roskilde Universitetsforlag

Krogstrup, Hanne Kathrine (1996): Brugerinddragelse i kvalitetsvurdering af sociale indsatsområder. Nordisk Socialt arbejde 2/1996. Scandinavian University Press; pp 114-129

Krogstrup, Hanne Kathrine (1997a): User Participation in Quality Assessment. Evaluation, pp 205-224, vol. 3 (2), Sage Publications

Krogstrup, Hanne Kathrine (1997b): Brugerinddragelse og organisatorisk læring i den sociale sektor; Denmark; Systime

Krogstrup, Hanne Kathrine og Jakob Tjalve (1999): Top-down og bottom-up orienteret brugerinddragelse I: Krogstrup, Hanne Kathrine: Det handicappede samfund - om brugerinddragelse og medborgerskab; Gylling; Systime

Krogstrup, Hanne Kathrine (2001): Brugerinddragelse i evaluering - top-down og bottom-up perspektiver I: Dahler-Larsen, Peter og Hanne Kathrine Krogstrup (red.): Tendenser i evaluering; Odense; Odense Universitets Forlag

Lipsky, Michael (1980): Street-Level Bureaucracy. Dilemmas of the Individual in Public Services; New York; Russel Sage Foundation

Mithaug, Dennis (1996): Fairness, Liberty and Empowerment Evaluation I: Fetterman, David M. Shakeh J. Kaftarian & Abraham Wandersman (eds.): Empowerment Evaluation. Knowledge and Tools for Self-Assesment & Accountability; Thousand Oaks; London, New Delhi; Sage Publications; pp 234-255

Otter, Carsten von (1985): Brukare og myndigheder I: Ingar Palmlund (red.): Utvärdering av offentlig verksamhet; Stockholm; Liber Förlag.

Patton, Michael Quinn (2001): Evaluatin, Knowledge Management, Best Practice, and High Quality Lessons Learned. American Journal of Evaluation, 22 (3) pp. 329-336.

Patton, Michael Quinn (1997): Toward Distinguishing Empowerment Evaluation and Placing it in a Larger Context. I: Evaluation Practice; 18 (2); 1997; pp 148-163

Rothstein, Bo (1994): Vad bör staten göra; Stockholm; SNS Förlag

Stake, Robert E. (2000): A Modest Commitment to the Promotion of Democracy: I: New directions for Evaluation; no 85; spring 2000

Stufflebeam, Daniel (2001): Evaluation Models I: New Directions for Evaluation. A Publication of the American Evaluation Association; number 89; Spring 2001 pp 7-106

3 EVALUATION MODELS AND METHODS

User evaluation in practice - the UPQA model

Hanne Kathrine Krogstrup

On a beautiful day in September 2003 I had the pleasure of presenting the UPQA model (User Participation in Quality Assessment. In Danish, the BIKVA-modellen for Brugerinddragelse i Kvalitetsvurdering) for a most sympathetic and zestful Finnish audience.

It is interesting that the model is being included with this article in the anthology Evaluation in The Nordic Countries, as it allows me the possibility of presenting it in writing to a wider Finnish audience.

The UPQA model assigns users of social services a unique role in evaluations. One of the members of the audience was doubtful of the possibility of including Finnish users in this specific evaluation context – users who were not accustomed to being involved due to a specific Finnish history and tradition. I have subsequently reflected upon this question and today I would ask the counter question: What comes first, the chicken or the egg? The UPQA model may be the tool that opens up the prospect of influence, and applying the model can, perhaps, be seen as a way of training staff, evaluators, and users in influence rather than influence being the precondition for applying the model. After all, the issue is one of dialectics, I guess.

Background

The UPQA model has been developed in response to growing demands to involve the users in the process of evaluation in the mid-1990s. The model available at that time was primarily user satisfaction surveys that were criticized for not actually evaluating user satisfaction. In most cases, these surveys assess satisfaction on the basis of questions/criteria defined by those other than the users and are not capable of 'capturing' the essence of what is important for the users. In addition, user satisfaction surveys are seldom appropriate for challenging existing social work and for producing new ideas for how to develop the services. However, it would be erroneous to see the UPQA model as the replacement of more traditional user satisfaction surveys. The two approaches have different potentials and should perhaps rather be viewed as supplemental.

UPQA is an abbreviation of User Participation in Quality Assessment. The model was originally developed within the realm of social psychology in Denmark and was published for the first time in 1996/1997 (Krogstrup 1996; 1997a and b). The model has subsequently been used for many different efforts inside and outside the Scandinavian countries. In Norway the UPQA model has been used for evaluating child welfare in a Sami municipality (Saus 2001), and in Scotland in connection with evaluating activation

of mentally disabled persons (NSF 1998). In a Danish context, the model has, for example, been used for evaluating day care centres in the municipality of Gilleleje.Græsted, achievements within the housing-social field in Horsens Municipality (Rubinke 2000), evaluation of social-psychic services and in connection with the planning of services for disabled citizens (Hansen et al. 1999) and in a series of other municipalities.

Furthermore, the model is referred to in, for example, the Swedish (Dahlberg & Vedung 2001), German (Kromrey 2001) and Italian (Bibliografia 1997) literature.

Object

The object of the UPQA model is, through the inclusion of users in evaluations, to secure a correlation between the public services and users' perception of problems, and hence between the users' perception of problems and the social work at different levels in the organization (Krogstrup 1996 and 1997a). The idea is that the users hold important knowledge that can contribute to the goal-directing of the public sector services. The evaluation process is bottom-up, oriented toward learning, and is expected to contribute to methodological development. The evaluation starts with the problems that the users find relevant, but also includes the front line staff (employees in direct contact with the users), managers, and politicians (Krogstrup 1996, 1997a and b).

The end is achieved by:

a) the users defining the problems to be evaluated and assessing the social work drawing on their own experiences b) relevant stakeholders (e.g. front line staff, managers and politicians) are confronted with the users' opinions, assess the problems raised and take a stand on the positive and negative criticism aired by the users c) continuous dialogue and development and making sure that the evaluation is followed up by action.

The Outlines of the UPQA Model

In the UPQA model, the user's assessment of social work is an important source for questioning and challenging practice. The following is an ideal typical presentation of the method of the model with a view to describing the principles governing the evaluation process:

The foundation of the reflexive process is that the users of the services in a group interview are asked to express and justify 'why they are satisfied or dissatisfied' with the services offered. It is important that questions are open-ended. The result of this group interview will be a series of experiences and subjective assessments of the social work, which are subsumed under thematic headings. From the user's perspective, quality is most often reflected in the relationship between users and social workers (Lipsky 1980). Therefore, users will often point to their experience with and assessment of this relationship (experience shows that 80-85% of the themes raised by the users concerns the relationship between users and front line staff).

A hypothetical example of a quality theme could be that users find it difficult to talk with social workers because they feel that the latter tend to be

the ones defining the problems in their lives and the solutions. In effect, the users feel deprived of personal initiative.

The user's the maticized assessments of the social work are presented to the social workers prior to a group interview during which they state their opinion of what could be the reason for the user's experiences. The social workers may, for example, point out that they find the users increasingly void of initiative and therefore see it as their duty to 'start something – to make something happen'. A discrepancy has been identified between user needs and social services. Furthermore, the front line staff may refer to the perception of the administrative management as continuously rejecting a call to implement user initiatives. A new discrepancy has been identified in the organization. The statements from the group interview with the social workers are now ideally presented to the administrative management and the politicians to get their response. Inclusion of the administration would be an illustrative example.

The users' and the social workers' observations are presented in an adapted form to the administrative management to get their assessment of the reasons in a group interview.

They may identify the politicians as key quality actors, and the latter are therefore presented with a summary of statements from users, front line staff and the administrative management in order to get their assessment. The politicians may pass the 'ball' to the social workers who are then requested to take a stand on the politicians' statement, etc. In this way the 'quest' continues for discrepancies related to the quality themes identified by the users (Krogstrup 1997a, 1997b).

The figure below illustrates the process of the UPQA model:

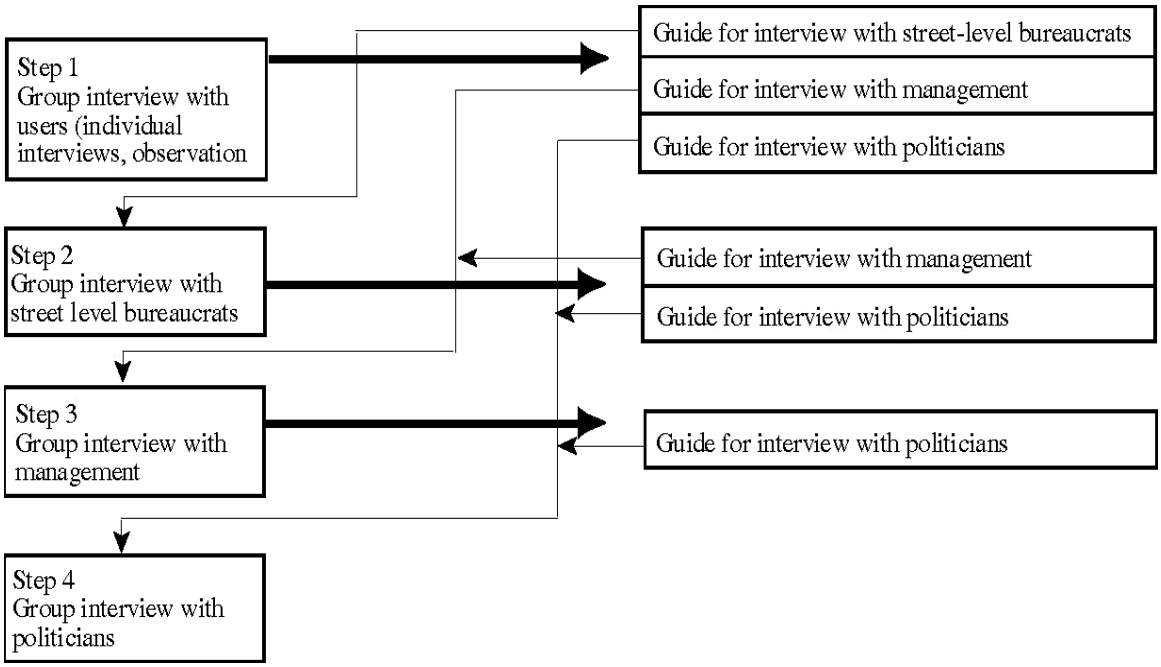

Figure 6. 1 The UPQA model, preparation Data processing Interview guides method

The boxes to the left show the stakeholders included in the evaluation and the applied method. The bold horizontal arrows illustrate the processing of interview data in preparing the interview guide for the 'next link in the chain'.

The 'thin' arrows illustrate for which stakeholders the interview guide is intended: Front line staff are confronted with user statements concerning their action norms. The administrative management is confronted with statements by users and front line staff relating to its action norms. Politicians are confronted with statements from users, front line staff, and management relating to their action norms.

Each step in process follows, in principle, the same recipe:

a) Selection of stakeholders for participation (users, front line staff, managers, politicians) b) Appointments and information about group interviews c) Group interviews with the stakeholders d) Data processing of group interviews e) Systematization of statements from interviews f) Processing of interview guide for the 'next link' in the chain.

From Qualitative to Quantitative

The UPQA model is based on qualitative date, the criteria for the evaluation are defined in the evaluation process, and the model is oriented toward learning. The qualitative group interviews can be used to qualify a subsequent survey.

First of all, the group interview facilitates categories in the survey that are founded in the users' interpretations of the services. This opens up categories and understandings that are unique in relation to the questions/criteria that non-users might formulate.

Second, the qualitative interviews help in the learning of the language of the users, which may help to qualify the way in which questions are formulated in a questionnaire. One of the greatest difficulties in designing questionnaires is to ensure that we get responses to what we wish to ask about. When respondents give answers that have nothing to do with the questions, the reason is most often that they respond to what they think they are being asked. The respondents answer on the basis of their different background, knowledge and experiences (Hansen & Andersen 2000:110). Therefore qualitative interviews might represent an important sources for asking exact questions that yield valid answers.

Third, preliminary qualitative interviews can help create support in the organization for implementing a survey (O'Brian 1993) that, for example, takes as its point of departure a user perspective.

A questionnaire can include several types of questions: 1) Factual questions about behaviour, events, external properties and characteristics; 2) Cognitive questions, the purpose of which is to measure the respondents' knowledge about certain issues; 3) Attitudinal questions, the purpose of which is to measure the respondents' views of or attitudes towards certain issues. Conducting a survey may prove a difficult process, but owing to a consideration of space, I cannot elaborate on the relevant problems and pitfalls. For a detailed theoretical description see Andersen & Hansen (2000).

In connection with the evaluation of kindergartens in a Danish municipality, the UPQA model has functioned as an integrated part of the municipal evaluation practice. In year n1 a UPQA model is conducted, in year n2 a survey is conducted designed on the basis of qualitative data from the UPQA model, in year n3 no evaluation is conducted. In year n4 a new UPQA process starts, etc.

The UPQA model can easily yield a substantial outcome without being linked to a survey, but the latter offers a unique possibility for including the users in the formulation of questions.

Experiences with the UPQA Model

During the period 2001-2002 the UPQA model has been tested in collaboration with a number of Danish municipalities. The most significant themes/problems proved to be to obtain the users' consent, choice of interview methods, the evaluator role in particular related to the group interview with the users, and conducting the interviews. The reservations of the front line staff toward including the users are a particular problem. Reservations are in part owing to considerations of the users but also to the front line staff's protection of themselves (Dahler-Larsen & Krogstrup 2003). However, results from the test municipalities show that when handled with care and consideration, it is possible to overcome these problems. The most serious barriers to implementing the UPQA model is assessed to be the reservations of the front line staff toward being confronted with their own norms for acting as interpreted by the users. This problem must be managed seriously in order to exploit optimally the learning potential of the UPQA model. It is important that potential reservations are discussed openly, that the front line staff are familiar with the fundamental idea of the UPQA model, and that the evaluator is aware of this problem in the evaluation process.

The recipe for the UPQA model recommends conducting group interviews with mangers and politicians about issues on which they have an influence. Another important experience from the test municipalities is that it might prove advantageous to include these links in the control chain in various ways. And this inclusion is important. Not primarily to enable managers and politicians to control front line staff, but because managers and politicians are, to a large extent, responsible for implementing the social politics and, equally important, their dispositions affect directly or indirectly the quality of the services supplied – as defined by the users. Hence, the model implicitly stresses that social work cannot be seen as solely concerning the relationship between the user and the social worker, and the evaluation must include the contextual conditions for this relationship.

The Role of the Users

The users are assigned the key role in the UPQA model as triggers of learning. The underlying raison d'être for including the users is that they possess a knowledge that can contribute to developing the public sector. Through

dialogue with the users, among others, the public organizations will get to know future demands (Ranson & Stewart 1994:176). According to Ranson and Stewart (1994:178) an important method for developing the public sector is to give, among others, marginalized groups an opportunity to communicate ideas and problems that would otherwise not be heard or only assigned secondary importance. Marginalized groups are thus assigned a change potential which is rooted in their criticism of the current practice of the public sector and in their ideas about future practice. By including the users, issues are rendered visible of which front line employees, managers, and politicians are unaware.

The Concept of User

In the UPQA model, conducting an open-ended interview with the users in the preliminary phase triggers learning. The purpose is to make the users describe their view of the social work in their own terms, interpretations, and dimensions. The other stakeholders are subsequently confronted with these descriptions. This method complies with the user concept of the UPQA model: Users are individuals receiving services from the public sector. The reason for including the users in the UPQA model is that their particular interests, preferences, and capacities should be respected. But the users are not only users, they are also citizens. The concept of user narrows down the rights of the users in relation to a) equality before the Law; b) political rights as voters; and c) social rights that ensure them a material safety net. But – and this is important – embedded in the concept of citizen are also duties: Users must subject themselves to control and the political agenda. Society has a general interest implying that its task is not only to consider the needs of the users. Legislation based on local political and financial priorities defines considerations of the general interest. The local municipal budgets define and prioritize the services offered to various groups within the social realm (Krogstrup 1997b:80). For a further discussion of the concepts of customer, client and user see Krogstrup 1997b:77-94.

The task of the public sector is thus not solely to cater for the particular interests of the users, the public sector must also consider the general interest. This entails that social work must necessarily be performed within the framework of political objectives, budgets, and legislation and what in a wider sense is perceived as socially legitimate.

This is the framework within which front line employees must perform social work. In directing tasks toward the users, front line employees have a certain amount of discretion. Furthermore, front line employees have a series of professional and ethical considerations to safeguard, and they must put the tasks into order of priority: For example, should one person in a group home be given the possibility for taking riding lessons or should the whole group go to the beach? At the institutional level, considerations of particular interests must always be balanced against the general interest. So the front line employees' task is not solely and at any time to meet the needs and demands of the users, but to consider the particular interests, capacities and preferences of the users to the widest extent possible within the institutional framework.

In other words, quality from the user perspective comes into existence in the meeting between front line staff and users. In 80-85% of the situations, user statements in the preliminary group interview concern the relationship between user and front line employee. It is in relation to the latter that the learning potential of the UPQA model is expected to be strongest, but with significant affects on other levels in the organization. The justification for beginning the UPQA model with an open-ended interview with the users is to generate data on how the users experience the institutional order.

Challenging the Institutional Order

Any institution has an institutional order. Actors create the institution order through establishing routines that define patterns of actions, which are repeated in everyday life and hence become meaningful to the involved actors. Habits and routines are, for example, necessary for the staff and users in a group home to avoid experiencing everyday life as chaotic (Berger & Luckmann 1966:173). Habits and routines serve a vital purpose in that they simultaneously control the behavioural patterns of the actors and set the bounds to acceptable behaviour. It is, for example, institutional practice in certain group homes for persons with disability that residents have keys to their rooms while in other institution this is inconceivable. The institutional order determines what is perceived as appropriate and inappropriate behaviour, which kinds of questions can be raised and which cannot, etc. Compared with the life of persons with disability in the large institutions, persons with disability living in group homes have far greater influence on their own lives – or the knowledge and values of the staff have changed. In defining the nature of good habits, routines, and 'appropriate' behaviour, the employees have the upper hand. The UPQA model challenges this power through the preliminary open-ended group interview with the users. This is perhaps the fundamental reason why some employees feel uncomfortable when the UPQA model is opted for as the point of departure for an evaluation – it might entail an actual development process.

The UPQA model rests on a knowledge argument that can be summarized as the users' knowledge/experiences contributing to goal-directed social work in regard to the implementation of political objectives. And by challenging the institutional order it contributes with input towards future development.

Inclusion of Front Line Staff, Managers, and Politicians

The argument for including front line staff as the next link in the 'interview chain' is that they have specific and legitimate considerations to safeguard, such as a) legislation and political decisions b) balancing general interests and considerations of the collective. In addition, the working conditions of the front line staff, including the culture, structure, and processes in the organizations in which they work, affect to a considerable degree the way in which they perform the tasks. The group interview with the front line staff will reveal their interpretation of how these considerations affect their task

performance and the interview is guided by a design that is in keeping with the outcome of the user interview.

Front line staff will often identify legislation, managers and politicians as important actors exerting influence on task performance. This is the argument for including managers and politicians. The recipe for the UPQA model emphasizes that managers and politicians are only confronted with statements by users and front line employees on whom they have an influence and can act in relation to. But managers and politicians often see the task performance of front line staff as their responsibility and they are therefore interested in the outcome of group interviews with users and front line staff.

Concluding Comments

This article may be too condensed. In spite of this, hopefully it functions as an appetizer. It presents the methodology of the model, describes concisely certain experiences with implementing the model, and the underlying theories. More detailed descriptions of implementing the UPQA model, the theory underlying the model, and experiences with applying it can be found in Dahler-Larsen and Krogstrup (2002): Nye Veje i Evaluering [New Avenues in Evaluation]. The publication is in Danish but is now available in Finnish (Asiakaslähtöinen arviointi Bikva-malli, FinSoc arviointiraportteja 1/2004)

References

Andersen, Bjarne Hjorth og Erik Jørgen Hansen (2000): Et sociologisk værktøj - Introduktion til kvantitativ metode. Hans Reitzels Forlag

Berger, Peter og Thomas Luckmann (1966): Den samfundsskabte virkelighed; Viborg 1992; Lindhardt og Ringhof

Bibliografia internazionale sulla valutazione (1997): Seval Bulletin nr. 4

Dahlberg, Magnus og EvertVedung (2001): Demokrati og brukerutvärdering; Lund; Studentlitteratur

Dahler-Larsen, Peter og Hanne Kathrine Krogstrup (2003): Nye Veje i Evaluering. Forlaget Systime Hansen, Hans Christian, Tove Røsvoll, Anne Skov og Bente Hansen Kermenoglou (1999): På besøg i konvolutternes Land; Frederikshavn; DAFOLO

Krogstrup, Hanne Kathrine (1996): Brugerinddragelse i kvalitetsvurdering af sociale indsatsområder. Nordisk Socialt arbejde 2/1996. Scandinavian University Press; pp 114-129

Krogstrup, Hanne Kathrine (1997a): User Participation in Quality Assessment. Evaluation, pp 205-224, vol. 3 (2), Sage Publications

Krogstrup, Hanne Kathrine (1997b): Brugerinddragelse og organisatorisk læring i den sociale sektor; Denmark; Systime

Kromrey, Helmut (2001): Evaluation - ein Vielschichtiges Konzept. Begriff und Methodik von Evaluierung und Evaluationsforschung. Empfelungen für die Praxis. I: Sozial Wissenschaften und Beruftspraxis (sub), 24. Jahrgang, Heft

Lipsky, Michael (1980): Street-Level Bureaucracy. Dilemmas of the Individual in Public Services; New York; Russel Sage Foundation

NSF (1998): The Ladder Initiative - An Evaluation of Employment Training and Support for People with Mental Health Problems; European Community - European Social Fund

O'Brian, Kerth (1993): Improving Survey Questionnaries Trough Focus Groups. Davis Morgan (Ed.): Successful Focus Groups. Thousand Oaks: Sage

Ranson, Stewart & John Stewart (1994): Management for The Public Domain - Enabling The Learning Society. St. Martins Press

Rubinke, Sanne (1999): Projekt Vesterled; Horsens Kommune

Saus, Merete (2001): Barnevern i en samisk Kommune - brukerinndragelse i kvalitetvurdering - BIKVA; Barnevernets utviklingssenter i Nord-Norge

Evaluation and Development: the Most Significant Change model

Riitta Seppänen-Järvelä

Introduction

The central concern of this paper is to introduce and explore a novel evaluation approach, the Most Significant Change (hereafter called MSC). Before focusing on this topic I will begin with an overview of the nature of renewal projects, especially in the context of organizational development. After that, I will introduce the idea of process evaluation as a framework for the MSC.

As a present day feature, traditional forms of governance have been gradually replaced with development-like practices stressing partnerships and interaction: Various renewal activities like programs, projects and pilots represent the new style of policy formulation and implementation. In this light, the support of a wide variety of development efforts has become an important element in the field of social welfare and health, both at micro and macro levels. Evaluation alike represents this improvement-oriented, managerial trend, though nevertheless, it is a natural element in all kinds of renewal activities.

It is obvious that we now live in an era of evaluation, especially in Europe. As Robert Stake has said, evaluation has been with us for a long time. And there is much more to come (2003:1). Since the 1960s, when a randomized experiment was almost the only preferred design, the evaluation field has changed a lot. In the course of time, the evaluation paradigm itself has broadened and become more pluralistic. Carol Weiss (1998) has noted that the increasing use of qualitative methods has been one of the more substantial trends. Another characteristic has been the rise of internal evaluation (Sonnichsen, 2000). An increased awareness of the advantages of qualitative methods and the criticism of black-and-white summative evaluation has pushed many evaluators to invent qualitative tools for practical evaluation tasks.

This was the general background when the MSC method was first developed by sociologist Rick Davies in Bangladesh in the early 1990s. At that time he worked in a large savings and credit project, where the direct beneficiaries were women. In that particular project he came face to face the problems of traditional outcome-oriented evaluation, something which prompted him to construct an alternative way of assessing developmental processes (Davies, 1998.) The awareness of methodological issues has especially increased among evaluation experts and practitioners in the field of international development (e.g. Oakley, 1990). The approach, which since then has been applied in several other contexts, also goes under names like "the narrative approach" or "story approach" (Davies, 1998). I myself became acquainted with this approach in a workshop in the Australasian Evaluation Society's conference held in Canberra 2001. Doctor Jessica Dart (who gained her experience of using MSC as an evaluation tool in a large-scale dairy extension project) facilitated a workshop at the conference where

the tool was demonstrated and discussed by the participants. To me, MSC sounded like a very promising alternative in terms of process evaluation and promoting organizational learning.

The nature of renewal projects

Nowadays, project work is the way to organize development activities. We even tend to see all kinds of developmental efforts as projects that are temporary, task-oriented, and serial. The idea of the 'project' is a powerful mental model that has entered into our personal and working life. According to traditional ways of thinking—alive and kicking in the field of project management—a project is a neutral tool without any assumptions or further implications (Packendorff, 1995). In this light, project work is above all a matter of project management, a growing professional area that has created its own qualifications and certifications (http://www.pmi.org). Nevertheless, it is wise to keep in mind that with regard to evaluation, the characteristics of project work have an influence on how a specific development activity will be carried out.

Contradictions within an organization are the typical drivers for development efforts. At the level of a work community, the meanings that individuals give to these contradictions vary a lot. In this sense, it is understandable that stakeholders and individuals have different and potentially conflicting interests and aims, and paying attention to creating a shared vision is an important though vulnerable issue. In addition, the fact that there are multiple interests and hidden agendas means a project is also an arena full of human interaction. The success of development work is dependent on such things as social capacity, creativity and trust. Considering these fragile elements, the key element is personal emotional experience that either nourishes one's involvement or not. Human experiences based on feelings are a crucial source of evaluation information that qualitative and interpretative methods have made possible to take into account (e.g. Stake, 2003: 6). The figure 7.1 below illustrates the wide variety of elements in human consciousness.

Development work starts from the premise that it is ideologically and normatively oriented. It struggles for social betterment. That requires personal and collective belief in change and the ability to be visionary. Also it is characteristic of development work that processes tend to be long-lasting and non-linear, and are more likely to be meandering and uncertain in their nature (Seppänen-Järvelä, 1999.) To my mind an evaluation approach that could promote developmental aims and improve interventions must be more process- than outcome-oriented, especially because measuring progress in learning and change is difficult: the payoffs are subtle and delayed, and there are more likely to be strategic advantages, such as greater flexibility and responsiveness (e.g. Brodtrick, 1998: 87). All this makes me wonder how well the project concept would fit the needs of development work.

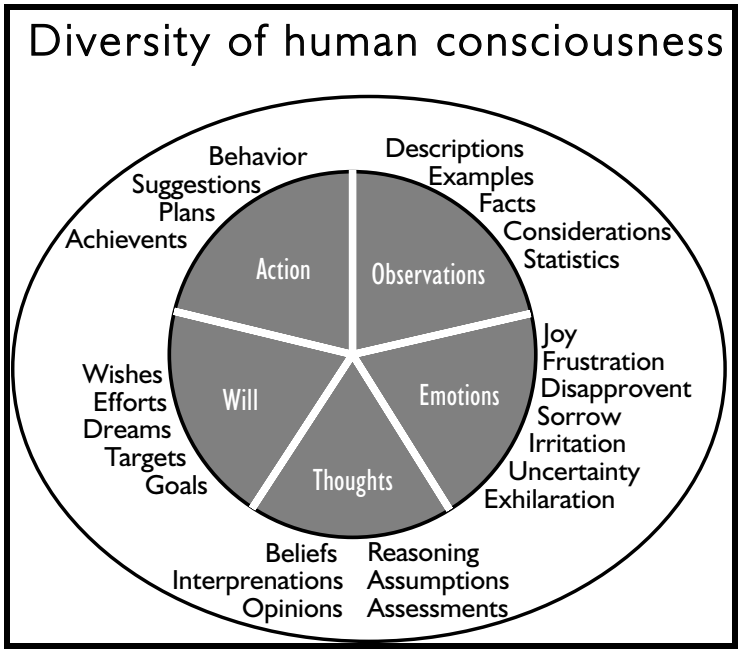

Figure 7. 1 The elements of human consciousness

Improvement through process evaluation

According to Weiss (1998: 4), evaluation is defined as a systematic assessment of the operation and/or the outcomes of a program or policy, compared to a set of explicit or implicit standards, as a means of contributing to the improvement of the program or policy. This last element of the key components refers to the developmental purpose of evaluation. Besides the outcomes that the original evaluation focuses on, evaluation questions will increasingly deal with the process of the program. So, whereas the primary use of traditional evaluation used to be judging merit or worth, process evaluation aims at improving the program. Understanding what is happening during the program is important not simply because of its developmental function, but also for understanding outcome data: what the outcomes are outcomes of (Weiss, 1998: 9-10).

When evaluation is an integrated part of a development effort, it involves a process use of evaluation data. According to Patton (1997), there are four distinct kinds of process use, in other words, being involved in an evaluation process can result in: (1) enhancement of shared understanding, (2) increase of participants' engagement, (3) support and reinforcement of the program through intervention-oriented evaluation, and (4) program or organizational development. Forss et al. (2002) have extended the categories of process use introduced by Patton by identifying five different ways of using an evaluation process: (1) learning to learn, (2) developing networks, (3) creating shared understanding, (4) strengthening the program and (5) boosting the morale. All the above mentioned types are highly relevant with regard to the MSC approach. They stress the importance of collective interpretations and

pinpoint the focal role of continuous utilization of evaluation data that are fundamental principles of the MSC

It is important to notice that different types of evaluations ask different questions and focus on different purposes. The typical questions asked within process-evaluation are of the following kind: What happens when the program is actually working? What do the participants experience in the program? What are the strengths and weaknesses of the day-to-day operations? Is the program changing? If so, how? Answers to these kinds of questions can provide crucial assistance in strengthening a program and organizational development. In addition, process information helps in understanding outcome data and effectiveness. (Robson, 2000: 63; Patton, 1997:194.) In order to get reliable answers to these questions, it is essential to use a wide variety of data and methods, although process evaluations are commonly based on qualitative data.

An overview of the Most Significant Change approach

"In universities,
people know
through studies.
In business and bureaucracies,
people know by reports.
In communities,
people know by stories"
(McKnight, 1987).

The approach can be described as a qualitative monitoring methodology, involving the collation of stories of change that emanate from the field, and systematic selection of the most significant of these stories (Davies, 1998).

Why stories? As a matter of fact, use of stories is not unheard of in the evaluation field or in qualitative research in general. Particularly in program evaluation, learning histories are used to document organizational and personal change experiences, so that project staff can reflect on what has or has not been working. Storytelling is a very old, yet powerful way for people in communities to share wisdom, experiences, humour and information. Actually, our modern world is full of stories, such as the stories that are told in newspapers, novels, on television, and on the Internet. These are the stories that we tend to listen to passively – without actively participating in them.

Stories help us to talk together about issues that matter and on subjects that are typically linked with change. Stories can also carry hard messages and they help us to question our assumptions. In addition, people will remember stories. In light of this, it is understandable that stories help connect people by dealing with issues that affect their lives and communities.

The MSC approach is strongly process and improvement oriented. So, whereas the primary use of traditional evaluation used to be for judging merit or worth, process evaluation is aimed at improving the program. Weiss (1998) has underlined the fact that understanding what is happening during a program is important not only because of its developmental function, but also for understanding outcome data: what the outcomes are outcomes of. In terms of evaluation, stories were collected in order to answer

the process evaluation questions like what happens when the project is actually underway? What do the participants experience in the project?

The participatory nature of the MSC has some implications on how data is collected and analyzed. Typically, external evaluators, specialists or senior levels of organization carry out an analysis of evaluation information. Normally field level workers do not analyze data, they just collect it, or rather they simply forward information up their hierarchies for others to analyze.

In this case, the staff not only collect information about events but they make their own evaluation of that information, according to their own local perspective (Davies, 1998).

The MSC in practice

The MSC evaluation runs throughout a project's life span. In fact, it does not even have to be a project: it can be any time-bound entity, like a training course. In a nutshell, there are four steps to conducting the MSC process, though the number of steps vary with regard to what level of detail they can be broken down to.

1. Determine domains of change
2. Implement a system to collect stories
3. Implement a system to evaluate stories
4. Monitor the process

Step 1. Determine domains of change

The domains of change that are going to be monitored have to be selected. The domains are loose categories used to distinguish different types of stories. It is important to notice that the domains of change are not precisely defined and they are left deliberately fuzzy. It is actually up to participants and staff as to what kind of meanings they give to change (Davies, 1998).

The selection of domains is a collective task. It can be done by a project team, a steering committee/advisory group or a group established for this. Because people can have very different opinions on what should be monitored, it can be tricky to get a consensus over 2 or 3 of the most important domains. To support the selection process, methods like DELPHI techniques or Fetterman's (2001) empowerment evaluation can be useful in prioritising competing opinions.

It is recommended that the number of domains be kept down to three. It is important that one domain is left without a title. For example in the original MSC-project, the domains were:

-Changes in people lives

-Changes in people's participation (Davies, 1998)

-Open title (Davies, 1998)

Step 2. Implementing

Who are the storytellers? The pool of storytellers can be wide or very focused; all involved actors are welcome or it can be defined as a sub-group. A central principle is that participants should be those who want to participate. Intrusion will make it likely that people will not co-operate, they might become defensive or evasive. (Robson, 2000: 25, 34). When naming some storytellers it is wise to keep in mind the very purpose of an evaluation task: For what purpose, and for whom is the evaluation data collected?

Evaluation, in general, is very much question oriented. Similarly, in the MSC approach, storytellers can be helped by posing a question like: "During the previous few months, in your opinion, what do you think was the most significant change that took place in terms of your professional development in the training course?" The question helps the storyteller to generate a story and recall past experiences.

Stories should be in written format, so spoken stories should be recorded and transcribed.

The respondents were asked to formulate their stories in two parts. The first part is descriptive: What happened? Where and when did it happen? Who was involved? The intention is that there should be enough information written down that an independent person can verify a reported event. The second part is explanatory: A respondent must explain why he or she thought the change was the most significant out of all the changes that took place in that time period. In particular, what difference did it already make, or will it make in the future (Davies, 1998)?

What is important is that significance is not expected in any absolute sense. Neither is there an expectation for it to objective. It would be very much concerned with subjective values and priorities. The process of sampling the most significant events is purposive rather than random; so therefore, the intention is not to screen for the average but for something unique and ultimate. (Davies, 1998.)

Step 3. Implement a system to evaluate stories

When the stories are collected, it is time to evaluate them. Filtering the stories is a collective, interpretative process: In a meeting, participants review all the stories and select one from each domain that represents the most significant change from their perspective. A meeting where the stories are selected may be partially facilitated, with the facilitator catalysing discussion by posing questions (Davies, 1998). Reflection, which is a specific element of evaluation (Preskill & Torres, 1999), plays a key role in these meetings. Basically reflection is an essential tool for learning from experiences: reflection is a natural and familiar process, often spontaneous, and, at times, outside an individual's awareness (Daudelin, 1999).

First of all, the titles of the stories are written on a white board under the respective domains and they are read out. The stories within one domain are reviewed and analyzed together. After the discussion, each participant is allowed one vote for one domain and a vote is done by a show of hands. The idea is to come to agreement as a group (Davies, 1998).

The most important thing is to explain why one story is particularly valuable or misleading. An evaluation group is required to document the stories they selected and what criteria were used. This information is fed back to the stakeholders; the selected stories will be sent to the next entity in the project hierarchy. Naturally, it depends on the scope of the project as to what kind of evaluation structure would be the most useful (Davies, 1998).

Step 4. Monitor the process

The last but not least of the steps is the monitoring of the process itself. That means fine-tuning the process of collecting and evaluating the stories and utilising the accumulated information. You might ask yourself evaluative questions like: Is the process running smoothly, if not, why not? Has the quality of the stories improved during the process? Are the project staff gaining more of a shared vision? Monitoring is constant in its nature but it can be organized by milestones or checkpoints.

It is especially important to pay attention to the utilisation of information. The risk is that the narratives prove to be interesting, but when they are collected for the evaluation purpose, they also need to be linked with the interventions and the improvements coming about as a result of the project. By monitoring the process you help to ensure a learning impact.

Conclusion

In this article, I have introduced the Most Significant Change approach, a participative monitoring tool. It belongs to the wider qualitative evaluation stream where it has obvious similarities with developmental evaluation, which is both user-focused and improvement-oriented (Patton, 1997: 104–106). Another strong connection exists to empowerment evaluation, which stresses an active role for participants in evaluating and steering interventions (Fetterman, 2001).

Openness and mutuality are the strengths of the MSC. It gives a channel for constructive feedback and interaction between project actors of different levels, but most of all, it gives a tool to analyze and evaluate a change process through meaningful emotional experiences. The MSC gives space for actors to reflect, to make sense of the complex changes that occur as a result of project intervention. Nevertheless, it is also wise to keep in mind that evaluation is not inherently positive (Patton, 1998: 227). Evaluation is always an intervention that has various "side effects".

The MSC is not an answer to all evaluation questions; it is not recommended as a stand-alone approach for assessing impact. Especially given that there is general agreement across the field of evaluation that all dimensions—summative and formative, as well as internal and external—are required in comprehensive evaluations, it is advisable to combine various methods in a specific evaluation design (e.g. Owen & Rogers, 1999; Stake, 2003).

Evaluation always has its practical utilitarian purpose—it is intended for use (e.g. Weiss 1998)—which means that it is good enough to answer the

program questions. Narratives can be rich but easily fragmented and lead to some sporadic conclusions. On the other hand, the data itself might be so overwhelming and arousing that there is the risk of being overcome it. In this light, the utilisation of gathered information is essential: what is really achieved or not in the project and what does that mean with regard to improving the project implementation. The potential risk at process use of evaluation – resided in all kinds of evaluation—is the tendency to provide status quo. To avoid this and to conduct a good MSC evaluation requires very sharp observation and a readiness to change course.

From a practical point of view, collecting stories and interpreting them is relatively time consuming, and to achieve the full potential of the MSC requires a reasonable capacity to analyze qualitative data and to manage evaluation practice. Confidentiality of information can be a delicate issue, and because the line between public and private information is fuzzy, it is essential to pay attention to how personal information is treated with respect to privacy. Since people tend to show the best side of themselves, it can be a challenge to get both good and bad news. The organizational culture can affect people's willingness to bare negative experiences and to create the possibility of a competitive aspect to the process. Regardless of some of the pitfalls of the MSC, it can be a fruitful mechanism for dialogue, and it can play a vital role in enhancing organizational development.

References

Davies, R. (1998). An evolutionary approach to facilitating organizational learning: An experiment by the Christian Commission for Development in Bangladesh. In Mosse, D.. Farrington, J., and Rew, A. (1998) Development as Process: Concepts and Methods for Working with Complexity. London: Routledge/ODI, pages 68−83; and in Impact Assessment and Project Appraisal. Vol. 16. No. 3, pp. 243−250.

Fetterman, D. (2001). Foundations of empowerment evaluation. Thousand Oaks: Sage Publications.

McKnight, J. L. (1987). Regenerating community. Social Policy. Vol.18 No. 3, pp. 54−58.

Brodrick, O. (1998). Organizational Learning and Innovation: Tools for Revitalizing Public Services. International Review of Administrative Sciences. Vol. 64 No. 1, pp. 83–96.

Forss, K., Rebien, C. and Jerker, C. (2002). Process Use of Evaluation. Evaluation. Vol. 8 No. 1, pp. 29–45.

Http://www.pmi.org. Printed on January 9, 2004.

Packendorff, J. (1995). Inquiring into the Temporary Organizations: New Directions for Project Management Research. Scandinavian Journal of Management. Vol. 11 No. 4, pp. 319−333.

Patton, M. (1997). Utilization-focused Evaluation. The New Century Text. Thousand Oaks: Sage Publications.

Patton, M. (1998). Discovering Process Use. Evaluation. Vol. 4 No. 2, pp. 225–233.

Oakley, P. (1990). The Evaluation of Social Development. In: Marsden & Oakley (eds.) Evaluating social development projects. Development guidelines 5. Oxford: Oxfam.

Owen, J.M. & P.J. Rogers, 1999: Program Evaluation. Forms and Approaches. London: Sage.

Robson, C. (2000). Small-Scale Evaluation. London: Sage Publications.

Stake, R. (2003). Standards-based & Responsive Evaluation. Thousand Oaks: Sage Publications.

Sonnichsen, R. (2000). High Impact Internal Evaluation. A Practitioner's Guide to Evaluating and Consulting Inside Organizations. Thousand Oaks: Sage Publications.

Weiss, C. (1998). Evaluation. Methods for Studying Programs and Policies. Second Edition. Upper Saddle River: Prentice Hall.

Polyphonic evaluation - how to use dialogues in organizational and network development

Vappu Karjalainen

Introduction

Polyphonic evaluation is a qualitative evaluation approach in multi-actor settings, especially useful in network contexts. The background to polyphonic evaluation lies in the development and research of our network team at Stakes. We have carried out many network projects over the past ten years (e.g., T. Arnkil & Eriksson & R.Arnkil 1999, 2001; Karjalainen & Lahti 2002) and have developed the network concept in social welfare, health care and employment services, specifically in the context of services for child care and rehabilitation. The evaluation dialogues have their origin in the development of anticipation dialogues by Tom Arnkil and Esa Eriksson. Some colleagues (R. Arnkil & Spangar 2003) also talk about emergent evaluation with the same meaning in mind.

Dialogue in evaluation is not a new issue. The concept of dialogue has meant several different things in evaluation and it has been conceptualized in different ways. It belongs to the tradition of stakeholder-based and responsive evaluation approaches. It is said that a dialogic approach is one way of revitalizing praxis and the moral-political life in society (Schwandt 2001). Our starting point has been the concentration on practices i.e. the development of network methods in which dialogism is the vital basis. Polyphonic evaluation is one application of this thinking.

The purpose of this article is to present the theoretical basis for polyphonic evaluation and to exemplify it in evaluating the Finnish One-Stop Shop programme (2002-2003). The focus is on describing the polyphonic evaluation method rather than the results and conclusions of the evaluation itself.

Why polyphonic evaluation?

We are living in a society that is experiencing rapid evolution and change. The concept of 'second modern' means a societal modernization characterized by individualization and globalisation (e.g. Beck & Beck-Gernsheim 2002). Traditional bonds (nuclear family, kin, community) are torn apart as the post-modern information society emerges. If the individualism of the Enlightenment was more about 'being individual', then the second modern individualism is about 'becoming individual' (Lash 2002). Individuals are facing a do-it-yourself life with a good deal of uncertainty and unpredictability. If healthy individualization is disturbed, the individual's problems often become complex. Welfare services and expert systems have to meet this complexity. This kind of situation calls for network structures and methods. What are needed are flexible multisector services transcending

sector boundaries and co-ordinating expertise of various kinds. Actually we are facing a paradigm shift from sector-based services to network services. The expertise of the second modern—the skill to master complexity and uncertainty—is embedded in a network of professionals and other stakeholders. The network method of joining resources and applying dialogism in communication are developmental challenges, and this current trend in service development needs to be evaluated; we need to look ahead and consider how best to approach to this development. Polyphonic evaluation is one such approach.

There is an enormous amount of information in every network context. One challenge for the professionals is mastering the rapid evolution in network relations and the information flow they are facing. What they need is reflective knowledge management, to process the overflow of information and learn from it. Organizations and networks need learning spaces so that professionals, managers etc. can 'translate' information into knowledge, action and joint activities. Evaluation can facilitate this process.

In 'classical' double-loop learning, the evaluation of project achievements leads to a questioning of the original goals and hypotheses on which the project was based. A learning process is initiated in which not only goals, but also practices are looked at critically, and adjusted or changed as necessary. To achieve double-loop learning, two elements are needed: a forum where stakeholders feel free to express their concerns, and an outside facilitator who can act as a catalyst in bringing out sensitive issues (Argyris & Schön 1978). Evaluation dialogues (dialogic sessions) are thought to be forums for promoting real-time learning from the processes being evaluated. In the network context, ex post evaluation in the traditional sense often comes too late to inform the developers of adjustments and strategies for changing circumstances. A real-time evaluation approach—in contrast to the linear approach—can integrate horizontal and vertical learning into the development path of the organizations and networks under evaluation.

One-Stop Shop programme - an example of polyphonic evaluation

The evaluation dialogues presented in this article were developed for the evaluation of the One-Stop Shop programme (2002-2003) in Finland. The goal of the programme was to develop integrated approaches to employment, welfare and health services by building joint service units for the long-term unemployed. The programme was an arrangement between the employment authorities (Ministry of Labour), the social welfare and health care authorities (Association of Finnish Local and Regional Authorities) and the Social Insurance Institution, and was aimed at combining their resources for the rehabilitation, activation and employment of the long-term unemployed. The programme was a two-year trial and it was carried out in 18 cities (25 separate one-stop shop projects). During the period 2004–2006, the one-stop shops will be transformed into service centres for the unemployed.

The objectives of the programme were both quantitative and qualitative:

1) to develop a joint service model of the employment, social welfare, health care and social insurance authorities for multi-agency unemployed clients at the local or regional level,

2) to ensure that a minimum of 60% of the long-term unemployed clients of the one-stop shop will find some kind of enduring solution to their situation,

3) to see that the number of unemployment assistance and minimum income recipients goes down in the project cities.

The evaluation model is described in figure 8.1 in a simplified form.

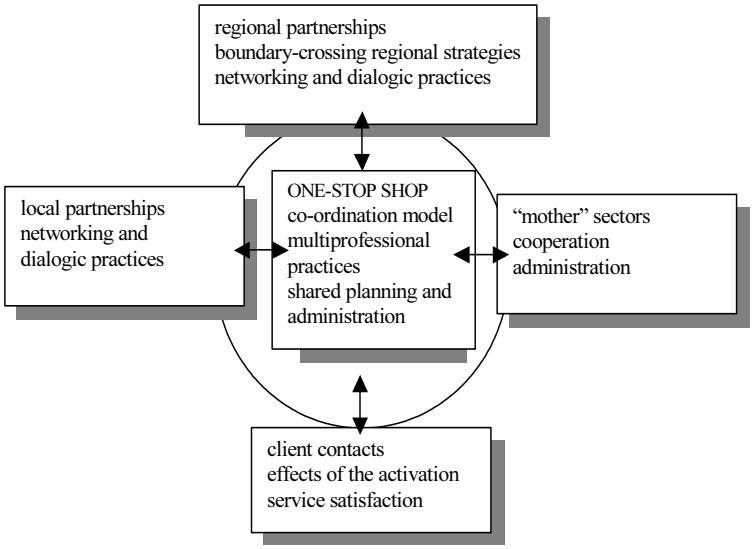

Figure 8. 1 The 360-degree evaluation model of the Finnish One-Stop Shop programme

The starting point for evaluation was to see each one-stop shop in its network relations at the local/regional level: What kind of position does the one-stop shop hold in the local network? The key dimensions—and themes for the evaluation dialogues—were internal functioning of the one-stop shop, client contacts, collaboration with 'mother' sectors (employment office, welfare office, health care centre etc.), co-ordination and collaboration with other horizontal partners in employment and activation (training centres, employers, third sector organisations etc.) and networking with the vertical partners, as with regional administration and strategy.

The evaluation task (objective 1) was to evaluate the collaboration and co-ordination of joint activities of the one-stop shops when developing employment, rehabilitation and activation of the long-term unemployed. Evaluation dialogues were developed for this purpose.

The quantitative objectives (2 and 3) of the programme - such as the effects of one-stop shop services at the individual level are evaluated by efficacy analysis (in the long-run). The experiences of the clients were evaluated both qualitatively (case analysis, client panels) and quantitatively (customer satisfaction analysis). The 360-degree model means that all the evaluation aspects are seen as a whole. The evaluators of various aspects

worked as a team[12] and in open dialogue with each other during the evaluation process. The evaluators have had their own learning loop, too.

The Ministry of Labour, which commissioned the evaluation, presupposed that the evaluation process would produce feedback information for local projects during the evaluation process. This motivated the evaluation team to pay special attention to learning loop arrangements and network learning.

What is polyphonic evaluation?

Polyphonic evaluation is formative evaluation. It has elements of empowerment evaluation (Fetterman et al.1996, 3-46) in fostering the improvement and self-determination of the actors involved. Polyphonic evaluation encourages the subjectivity of the participants by giving space and voice to each stakeholder. Evaluation dialogue, in fact, is based on subjectivity. There is no objective picture of a situation; on the contrary, each participant has his/her perspective in the observing system.

Polyphonic evaluation has been developed especially for multi-stakeholder and network settings when the communication and dialogue between various stakeholders is important but often difficult to arrange in a proper way. The idea is to facilitate dialogue and interaction within the innovative project network in order to promote self-evaluation and generate fresh knowledge for developing the network. Innovation is a network phenomenon and innovative practices are developed in interactive settings (Miettinen et al. 1999). Real-time evaluation can encourage this process. Polyphonic evaluation is a method of doing justice to the multi-dimensional network phenomenon.

Based on the network approach, the language and the dialogue—speaking and listening in a special way—are essential elements in polyphonic evaluation. The arrangement of dialogic forums is a method of bringing various voices to reflective discussion. The dialogic session is a kind of reflective workshop—a learning space—in promoting mutual understanding and network learning.

Polyphonic evaluation challenges the implementation of top-down programmes in the welfare society. In every programme, the key subjects of implementation and development are local, grass-root actors. How they see motivation, resources, opportunities, strategies etc. are basic factors when new practices are being implemented. As a bottom-up approach, polyphonic evaluation can facilitate collective reflection and revise the strategies and even the objectives of the programme.

[12] The members of the evaluation team: Timo Spangar, Robert Arnkil, Sanna-Mari Lytinen (Social Development Co), Vappu Karjalainen, Tuukka Lahti (STAKES) and Simo Aho (Work Research Centre, Tampere University).

Methodologically

The basic belief system behind polyphonic evaluation is hermeneutic (Cuba & Lincoln 1989). The purpose of the evaluation dialogues is to identify and describe various constructions the participants have concerning certain themes and bring those constructions in touch. The nature of this kind of information is interactive. The outcome is relational-responsive understanding of our being in the world. So the challenge is to 'know from within' (Shotter, 1993). At its best, an evaluation dialogue produces contextual understanding of the developmental phases of the project to be evaluated. The roots of the evaluation dialogues are in social constructionism.

A network consists of various actors, none of which has the power to determine the opinions, strategies etc. of the others. The network has no centre, the participants are linked to each other horizontally and/or vertically (see Castells 2000). Actors in the network have their own unique point of view instead of representing the overall picture. The challenge in a network is to encourage each participant to elaborate his or her own point of view. When one person is listening to another person, he or she is receiving impressions and beginning to elaborate his or her inner dialogues (Bahtin 1991). That person is 'looking ahead' to respond to the other person's words. So listening is an essential part of communication. This kind of a reflective discussion is the basis for understanding one's own position and that of others within a network. The principles of network thinking and dialogism are basic to polyphonic evaluation.

In the evaluation dialogues, solution-centred orientation is applied (de Shazer 1991) to break the conventional linear approach. The evaluation questions are formulated so that the present development of the project is approached from a 'solved' perspective. The future perspective creates an optimistic and innovative framework to tackle present problems and difficulties. The approach contributes to network learning and reveals the innovative resources with new subjectivities.

Polyphonic evaluation in practice - evaluation dialogues

Arrangement of the dialogic evaluation sessions

Each dialogic evaluation session is led by a couple of facilitators. The facilitators are external to the project but they are familiar with the phenomenon to be discussed and reflected on. One facilitator is in the role of interviewer while the other is in the role of a reflective partner and note-taker. In this case, the team leader of the one-stop shop invites the representatives of various sectors and other network-partners to an evaluation session (representatives are changed according to the theme of the dialogue).

A group of invited representatives (max 6-7) sit around a table at the front of the room and one facilitator joins them. If there are more representatives and e.g. professionals from the one-stop shop present, they form the audience and listen to the conversation. The facilitator asks each representative in turn the same evaluation question (see the questions on the following page) while the others listen. The next question proceeds in the same way.

Speaking and listening are separated. The questions can be shown on a screen so that all the participants in the evaluation session can see them. The other facilitator writes down the key views of the representatives on a flap chart or the notes are written on screen with an on-line computer so that they can be seen, read and corrected together if necessary (validation of the material). Finally, the facilitator asks the audience to reflect on their experiences of the theme discussed. The notes are stored in Word format and sent to each participant at the session and to STAKES as part of the nationwide evaluation material.

Training the facilitators

The facilitators have to know the principles of the dialogic approach in leading a dialogic session. There were 25 one-stop shops (in 18 cities) in the programme, and each one-stop shop and its network could have six evaluation dialogues during the two-year experiment. The facilitators should be available at all evaluation dialogues. So STAKES arranged training (5 days in all) for professionals from social welfare and the employment services who were interested in facilitating dialogues. The training programme was concerned with network reasoning, principles of dialogism, anticipation and future orientation and practical tips for facilitation of a dialogic session. The collective meeting days between six evaluation dialogues were worthwhile for reflecting on the experiences and impressions received by the facilitators when leading the sessions. The short training appeared to be enough and the facilitators managed to lead the dialogic sessions surprisingly well.

Sets of evaluative questions

During the two-year experiment, six evaluation dialogues were arranged with the one-stop shops and their networks. The dialogues had various themes according to the developmental phase of the experiment; the theme of the first dialogue was the basic tasks of the one-stop shop team, while the last concerned the dissemination of new practices and the transition from one-stop shops to service centres for the unemployed. The questions were discussed and formulated together during training days. The evaluation dialogues were:

1) Dialogue on basic tasks was an 'inside' session with the one-stop shop team. The purpose was a) to evaluate the basic tasks of the professionals and their relations with each other, and b) to promote teamwork and dialogue inside the one-stop shop.

Questions:

1. What kind of task do you have in the one-stop shop and how does it differ from your former (sector) work?
2. What have you learnt from your colleagues in the one-stop shop (from a social worker/employment counsellor/public health nurse etc.)? What have you learnt from the clients?

3. What would you give preference to in your task while working in the one-stop shop?
4. What kind of support do you need for that? What are you going to do to get that support?

2) The first dialogue with the external network was an anticipation dialogue and it was arranged with the local network-partners of the one-stop shop, mostly with the local/regional management group of the local experiment. The aim was a) to evaluate local/regional development of the joint service model, its positive and problematic features, and b) to invite network creativity and strengthen the commitment of the local administration and the leadership for the trial.

Future dialogue: The facilitator asks the participants to imagine that 'we all have moved into the future, one year ahead'.A year has passed. Today it is 15th January 2004. Your one-stop shop project has just come to the end. It was a big success. You have developed a well-designed model of joint service and obtained good results in activation and employment. The local actors acknowledge your activities.

Questions:
1. What makes you particularly happy about the outcomes of the project from your point of view?
2. What was your contribution to this positive development?
3. Were you worried about something a year ago when the second year of the project started and what helped to lessen your worries?

3) The dialogue on the development of collaboration was an evaluation dialogue between the workers in the one-stop shop and those in the 'mother sectors' (social welfare office, employment office, health care centre, local unit of social insurance institution). The purpose was a) to evaluate the client-based collaboration and network methods of the one-stop shop, and b) to encourage local partners in developing network methods and further networking.

Questions:
1. What are the good collaboration practices of the one-stop shop? With whom and in what kind of issues?
2. What are you worried about concerning the development of collaboration practices of the one-stop shop?
3. What would be the next step to lessen these worries? What could you do for your part?

4) The second dialogue with the external network was a continuation of the first external dialogue (see dialogue 2). It was arranged with the same participants about six months later and the starting points for the dialogue were the conclusions from the earlier session. The aim was:

a) to evaluate the evolution of the joint service model,
and
b) to encourage the management group of the local experiment to give further support to developing the one-stop shop. The notes on the first dialogue with the external network form the basis for this session.

Questions:

1. To what extent - since our previous session - has the bright future of the joint service come true from the point of view of your organization?

   Use the rating scale:

   - = setback

   0 = no progress

   + = some steps in a positive direction

   ++ = joint service is on a good development path

2. Which actions have especially encouraged this development?

3. What could you do during the latter period of the project to contribute to the bright future of the joint service?

5) Peer evaluation was arranged between two one-stop shops and their networks. The purpose was a) to evaluate two different models of joint service and generalization of practices to other contexts, and b) to promote horizontal learning and build mutual contacts between one-stop shops. Two one-stop shop teams (A and B) and their networks.

Questions to teams A and B:

1. What kind of lessons (DO's & DON'T's) would you like to pass on to a one-stop shop just starting to organize its activities:

   a) Concerning joint planning?

   b) Concerning joint client service?

   c) Concerning joint administration and management?

2. What impressions do the experiences of team A arouse in team B andvice versa? Which lessons were particularly interesting?

3. In the light of your experiences, which practices (see question 1) could be generalized to other contexts?

6) The dialogue on the transfer and dissemination of new practices to the service centre for the unemployed was the last dialogue. The one-stop-shop team, the management group and other network partners could participate in the session. The one-stop shops will become permanent service centres for the unemployed in 2004-2006. So the aim was a) to evaluate the developmental results of the one-stop shop and the possible dissemination of new practices in the service centres for the unemployed, and b) to encourage the local/regional network in further activity and co-ordination during the transition phase.

Questions:

1. What are the salient challenges from your point of view during the transition phase (from one-stop shop concept to the service centre for the unemployed)?

2. Which new practices developed during the one-stop shop experiment are especially suitable for and easily disseminated in the service centre for the unemployed?

3. What kind of new practices do you wish to be developed in the the service centre for the unemployed?

4. What kind of support do you anticipate the team at the service centre for the unemployed will need for developing the centre? What are you yourself ready to do for this development?

Experiences

From the standpoint of our evaluation team, the evaluation dialogues were successful. Each one-stop shop participated in at least one dialogic session, many in almost all. Some dialogues (e.g. basic task and peer dialogue) were more popular than others. On the whole, our evaluation team received good horizontal evaluation material for the bottom-up analysis of the One-Stop Shop programme. The notes on the dialogic sessions were carefully written and sent to the evaluation team in addition to the local participants in the dialogues. The results and conclusions of the qualitative analysis are reported in the interim evaluation report (Spangar et.al. 2003). The report has turned out to be useful for the local/regional developers and for the central administration in building up to the transition from one-stop shops to service centres for the unemployed.

The experiences of the facilitators were collected in reflective meetings between the dialogues. The beginning of the evaluation dialogues were somewhat confusing for the facilitators, the theoretical background for dialogism in the training was quite 'packed' and there was not enough time to practice leading a dialogic session. However, the experiences from the first session were promising; the local actors were interested in the new method and the theme of the dialogue (the basic tasks) was relevant. The dialogues with the management group (first and second external) were seen as not only challenging but also very interesting because of the emergence of a wider perspective on local activities to do with social and employment policy. The problem was that many executive-level participants were busy and were unable to come to the session. The facilitators' experiences were mainly positive and some became further inspired by dialogism and its applications, but there were also feelings of tiredness towards the end of the dialogues.

The biggest problems that the facilitators experienced concerned the termination of a dialogic session; how to make use of polyphonic discussion and the emergence of the various points of view from a longer-term perspective, how to facilitate the participants at the session to plan the next step in the development and co-ordination of local/regional activities. The concern deals with network learning and the re-adjustment of developmental activities. A dialogic session is an intervention in the development of network activities in a special way; it is reflective, innovative and future oriented. In a dialogic session the need for new network methods, collaborative practices, opportunities for rehabilitation, etc. can emerge but it is not a setting for deciding on how to develop those activities. The session it not a decision-making forum. What can be done in a dialogic session is to produce a plan concerning the further steps to be taken in promoting development (e.g. agree on a meeting with the management group in order to obtain support for the plan). So the role of the facilitators is to promote dialogue, joint learning and innovative activities. The evaluation dialogues are reflective and development-oriented network forums and learning spaces, but decisions concerning resources and other practical issues related to the project are taken elsewhere.

What about the experiences of the one-stop shops? A small Internet inquiry with questions on opinions concerning the implementation of evaluation dialogues was sent to the team leaders of the one-stop shops. According to the results, 82% conceived the evaluation dialogues as useful (3-4) and 12% as very useful (5) (scale 1-5). In one case, the one-stop shop saw no use for them at all. Evaluation dialogues were conceived as most useful in the self-evaluation of the local team and network activities. Dialogues also promoted development work and dialogic practices. Peer dialogue was most appreciated (70% favoured that). Learning from each other and networking with other one-stop shops was very inspiring. Local actors became more aware of dialogic practices and horizontal learning during dialogues but it is not possible to say how great their interest would be in applying dialogic methods to further contexts.

Discussion

Evaluation dialogues appear to be promising. Polyphonic evaluation is especially suitable

1) in multi-stakeholder (network) and partnership-building settings - in polyphonic phenomena - where the development of joint activities and network methods are the current challenge

2) in development projects/contexts of rapid change when learning and real-time adaptation to change is a challenge, and

3) when dissemination and the generalization of the outcomes of the project and collective learning spaces are prioritized.

The interest of professionals in methods and service structures that promise ways of getting out of complex situations is increasing. Correspondingly, network structures and cooperation over sector boundaries is becoming more and more general. Developments are rapid and we know rather little about how to manage the huge amount of information that is available in a network. What we need is an evaluation approach suitable for this type of multi-voiced and changing circumstances. Both the professionals and the evaluators have to master all the information within a network, make use of it and process it further via reflective forums into collective knowledge and wisdom. Our experience is that evaluation dialogues and reflective network sessions can give a good start for double-loop learning.

We have applied fairly structured evaluation dialogues and used external facilitators. In our experience, the facilitators at the dialogic session are necessary. It is important that they are familiar with the context of the phenomenon to be discussed and that they do not have any position of their own in the context to be reflected. Their task is to facilitate openness and subjectivity and they have to tolerate a good deal of diversity and uncertainty.

Our experiences of the pre-structured questions in the dialogic sessions are somewhat ambivalent. Some kind of structure in the session is necessary (such as the separation of speaking and listening), but obviously there could be more flexibility in the questions to be discussed. Additional questions which clarify or deepen the theme are necessary in the event that the discussion does not progress or remains at a very general level. The principles of

dialogism are openness and polyphony. It is worth considering how to apply these principles best in various contexts. This is a further development task in the facilitation of the dialogues.

The clients, in our case the long-term unemployed, have not participated in the evaluation dialogues (although their organisations have). In order to listen to the voices of the unemployed we have used client panels. In panels, the unemployed jointly reflect their own experiences of one-stop shop services. We took the view that participating in a dialogic session among the professionals would be unnecessarily demanding and stressful for the clients. Some 'personalities' would probably like to participate but the voice of the clients would in any case be in the minority compared with the number of professionals present. In the future development of the evaluation dialogues, it would be interesting to experiment with new kinds of forums to try and bring the clients into dialogue with the professionals. An arrangement worth experimenting with would be 'a two-table dialogic session': one for clients and one for professionals. The facilitators would interview each table in turn so that both 'parties' would listen to each other and get their own impressions.

In the evaluation of a project or a programme, the development of network and co-ordination activities are generally one part of the evaluation task and one approach to it. Other aspects, such as the efficacy of interventions and cost-benefit analysis are just as important, although they assume a different evaluation research approach. One interesting challenge would be to unite different evaluation aspects in one evaluation: How would it be possible to get the hermeneutic and the positivist approaches to cooperate and generate theoretically and contextually rich analysis within the same evaluation? This presupposes that the individual researchers within the evaluation team can have dialogue with others, respect expertise of a different type and possess the ability to cross scientific boundaries in building a collective and participatory understanding of complex social phenomena. A balanced methodology is important in rapidly changing evaluation contexts.

References

Argyris, C., Schön, D. (1978): Organizational learning: A theory of action perspective. Reading. MA: Addison-Wesley.

Arnkil, T., Eriksson, E., Arnkil, R (1999): Anticipation dialogues.Vertical and horizontal dialogue methods for strategic management and smart networking. Stakes, Themes 3.

Arnkil, T., Eriksson, E., Arnkil, R. (2001): Dialogisk utveckling av kommunernas service. Från sektorcentrering och projektkaos till flexibla nätverk. Stakes, Rapporter 260

Arnkil, R., Spangar,T. (2003): Evaluation, dialogue and learning in multi-stakeholder settings. Paper presented at the Fifth European Conference on Evaluation of the Structural Funds, Budapest 26-27 June 2003.

Bahtin, M. (1991): Dostojevskin poetiikan ongelmia. Kustannus Oy Orient Express.

Beck, U., Beck-Gernsheim, E. (2002): Individualization. Institutionalized individualism and its social and political consequences.Sage Publications.

Castells, M. (2000) Materials for an exploratory theory of the network society. British Journal of Sociology 5: 5-25.

deShazer, S. (1991): Putting difference to work. New York. Norton.

Fetterman, D., M., Kaftarian, S., J., Wandersman, A. (eds.) 1996: Empowerment evaluation. Knowledge and tools for self-assessment & accountability. Sage Publications.

Guba, E., G., Lincoln, Y., S.(1989): Fourth generation evaluation. Sage Publications.

Karjalainen, V., Lahti, T. (2001): Kokeilusta käytännöksi. Miten edistää pitkäaikaistyöttömien aktiivikäytäntöjen yleistymistä? Stakes, Raportteja 264. Saarijärvi.

Lash, S. (2002): Individualization in a non-linear mode. Foreword in Beck, U., Beck-Gernsheim, E. (2002) Individualization. Institutionalized individualism and its social and political consequences. Sage Publications. pp. vii-xiii.

Miettinen, R., Lehenkari, J., Hasu, M., Hyvönen, J. (1999): Osaaminen ja uuden luominen innovaatioverkoissa. Tutkimus kuudesta suomalaisesta innovaatiosta. Suomen itsenäisyyden juhlarahasto, Sitra 226. Helsinki.

Schwandt, T., A. (2001): A postscript on thinking about dialogue. Evaluation.Vol 7(2): 264-276. Sage Publications.

Shotter, J. (1993): Conversational realities. Constructing life through language. London. Sage Publications.

Spangar, T., Arnkil, R., Lyytinen, S-M., Karjalainen, V., Lahti, T., Aho, S. (2003):

Yhteinen palvelu kehittyy - kehittyykö yhteinen johtaminen? Työhallinnon, kuntien ja Kelan yhteispalvelukokeilun arvioinnin väliraportti. Työhallinnon julkaisu 331. Työministeriö. Helsinki.

Integrating Evaluation in Welfare Praxis – the Integrated Model

Göran Sandell

Ever since the early 1960s when Thomas S. Kuhn in The Structure of Scientific Revolutions defined 'paradigm' as "universally acknowledged scientific works which for a period of time provide a scientific community with model problems and model solutions", the concept of paradigm has been important in the debate over scientific theory. Kuhn makes a distinction between "normal science", when the majority of researchers within a discipline are agreed on the values, research problems, methodological questions etc. and periods of "revolutionary science", when a new approach questions the prevailing one and new paradigms and models can emerge. Analogous to Kuhn's general definition, the paradigm for evaluation in welfare can be defined as follows: "norms and ideas of what can and should be evaluated in welfare as well as how this evaluation can be conducted" (Sandell 1985–2003).

The what question comprises what is considered interesting as a research or evaluation problem, what can be studied or evaluated in principle and be within reach of scientific analysis as well as what is the aim and cognitive interest of evaluation. The how question contains everything from all-embracing cognitive theory questions of what is meant by knowledge, where it comes from and how it is validated, through the ways whereby these problems can essentially be approached (logical empirical / hermeneutic / critical, quantitative / qualitative etc.) to the concrete methods (experiment / observation / interview, etc.) of collecting, aggregating, processing and analysing the data.

Standpoints on these paradigm questions are in turn based on ontological basic values and outlook on humanity as well as on a research theory and welfare policy understanding and will eventually be crucial to which model is to be advocated in the evaluation of various welfare organizations and activities.

Evaluation in welfare systems calls for a new model within a new paradigm

Within the traditional paradigms, such as logical empiricism, hermeneutics and critical theory, and also within newer 'paradigms', such as realistic evaluation and attempts in postmodern feminist, constructivist, language systemic and narrative approaches, there are a range of methods that can be excellent in specific and restricted contexts. Nonetheless, models that can evaluate large-scale and continuous activities in welfare, such as social work done in an entire city, are lacking both in the Nordic countries and elsewhere. At the same time, the development of welfare quality requires infor-

mation about how the activities benefit the service users' lives. Continuous evaluation of user benefits should therefore be integrated into regular activities in social services, primary and elderly care, rehabilitation, psychiatry, etc.

Such an integrated evaluation needs a new paradigm with a scientifically congruent and feasible model, which can continuously evaluate the entire multidimensional activities by the welfare organizations - and all their users, contributions and results. For such an evaluation to work in practice, it is both ideologically fundamental and necessary for real implementation to develop a model which:

- is congruent with the basic values and ethics in the existing welfare legislation
- represents a fundamentally respectful and humane attitude towards the users concerned
- has a well thought-out view on what is meant by user needs and activity results as well as
  how this information can be obtained in valid, reliable and secure forms
- supports the ongoing welfare work and facilitates the care and treatment relations
  established between the professionals and clients
- examines the clients' life situation, needs for change and resources, various interventions,
  actors of change, as well as the results of these activities, such as how they benefit the clients
- carries out compilation, aggregation and analysis of information within reasonable costs and timeframes for regular activities in welfare organizations without the need for overlong
  external expert involvement
- can give continuous feedback on the received information and conclusions at all levels
- concerned without complicated cost- and time-consuming special analyses
- can at the same time function as a knowledge-base for research, R&D and statistics.

In order to meet all these requirements, there is in fact no other realistic way – from the ethical, legal, political, strategic, integrity-related, research-methodology, practical or financial perspectives – but to integrate evaluation in its entirety into the framework of comprehensive democratically steered welfare organizations and activities. To make this possible, confidence is in turn needed in the fact that welfare professionals can carry out their services together with the clients and that evaluation information can be gathered in mutual dialogue between personnel and users within the framework of this continuous welfare work. In order to cope with integrity, capacity, working hours and costs, this evaluation also needs tailored IT support enabling documentation, aggregation and result analysis in welfare praxis.

With this fundamental values and ambitions as a starting point, a model for integrated evaluation has been developed by Göran Sandell, through research and R&D projects conducted together with a number of different

welfare organizations and activities during a development process spanning two decades, and described in books and reports.

An advanced, user-friendly, time-reducing, cost-effective and internationally unique IT-support to the integrated evaluation model has also been created (by Sandell in a joint-venture cooperation with TietoEnator.

The ideology and principal structure of the model

Evaluation is conducted congruent with the aims, ethics and outlook on humanity present in welfare. The model takes the fundamental values of Nordic welfare legislation for granted – democracy, solidarity, comprehensive view, sense of togetherness, integrity and the reinforcement of one's own resources, self-determination and personal responsibility. Evaluation focuses on the life situation and needs of the clients, examining whether the activities can help them to meet their needs for care or change in various spheres of life. Knowledge about the results and benefits for the recipients is considered paramount and necessary in order to understand and assess the inputs, costs and quality of welfare.

The model is integrated into the regular welfare activities. Evaluation is incorporated in the activities and accountabilities of social welfare organisations, adapted to the tasks and target groups of the organizations and conducted within the general framework of welfare activities and democratic steering. The professional staff carry out information gathering for evaluation integrated into their ordinary work in mutual information-seeking structured dialogue-interviu with the people concerned. This dialogue can also contribute to an increased understanding and smoother co-operation in other investigative and helping efforts and procedures.

The information obtained – interview data, results, analyses and conclusions – are 'owned' by the "evaluative-welfare-organization", which can flexibly select the target groups, case status, and evaluation period for its own analyses. Evaluation can also be included in yearly activity reporting, so that information about people's life situation and results from the welfare-activities are taken into account in programme work, prioritising and planning.

The model represents a welfare-integrated actor paradigm, complete with a cognitive interest of development, evaluating welfare-activities in a continuous process by analysing the needs, contributions, results and patterns. In an ambition to create a learning organisation with continuous quality development benefiting the clients, feedback on the information derived from evaluation is channelled to the users, personnel, leadership and politicians to reflect on the benefits and total-quality in relation to the overall aims set for the welfare organizations.

Integrated evaluation has also developed logical, safe, integrity-protected and user-friendly

IT-support that facilitates time-saving documentation of the dialogic interviews and enables flexible selection of target groups, evaluation periods and analytic perspectives as well as facilitates simple and quick aggregation, as the results are combined with graphic presentations of the outcomes through predesigned diagrams.

Principles of an integrated analysis of needs and results

Evaluation in welfare organizations and activities is basically about examining whether the contributions made have done some good in the users' lives by meeting their needs for change and helping improve their resources and life-situation. People's needs and results from welfare-activities are, however, intricate, complex and multidimensional with a reciprocal impact and therefore integrated evaluation has developed the following research principles:

Information about the life situation and needs for change is sought in a dialogic form between the users and professionals, whereby evaluation tries to take an overall view by studying the clients' situation and needs in several spheres of life that are divided at structural, interpersonal and individual levels.

Needs analysis means that you examine not only the shortcomings in actual life situations but also people's own resources to meet their needs for change. These needs are identified and documented in terms of 'aims for desired change' within a given sphere of life – so as not to confuse these real needs in the lives of the clients, with wishes or demand for different interventions. Studying whether the clients have developed resources to 'live in a need-satisfying way' does not mean that the needs cease to exist but that they now can be met with the resources of the clients, their families and social network, without welfare interventions.

The results are defined as the impacts of welfare-actions on the clients' internal and external life-situation. If the needs for change are expressed as 'targets for desired change', it is also possible to examine whether these targets have been met. Thus, the results should not be confused with the means of activity; interventions, management, budget, performance, staffing etc. - when the results are not primarily about what happens with the organization, personnel and money – but what happens in the lives of the people in question.

Impacts on people's lives are studied in many spheres of life in pursuit of an overall view enhancing the security and validity of information. Follow-ups are conducted as analyses of ´time series´ integrated into ordinary work and in dialogue between professionals and clients. The appropriate time for follow-ups is recommended every 6 month but is finally decided by the local activity, as the time perspective is governed by the goals, tasks, law, policy and the clients' life situation. Knowledge of the results' durability often requires follow-ups also after the interventions and contributions in a given case have been discontinued.

Evaluation should not be restricted to isolated interventions or programmes. Instead it should encompass as many actually implemented interventions as possible as well as include the users' own resources and initiative, as well as commitment on the part of the relatives and close persons. Integration into the ordinary welfare systems includes all the users with all their shortcomings and services in the entire population evaluated - a fact justifying fundamental caution against drawing generalised conclusions from too small target groups.

The model's IT-support allows making totally flexible choices of target-groups out of the compiled study population through selection criteria derived from both the people's background conditions and interventions made, as well as a flexible result analysis concerning the evaluation period and follow-up history as well as whether the target group includes cases with ongoing, completed or discontinued contributions and interventions. Flexible selection of target-group, case-status, evaluation-period and result-dimension in turn make it possible to:

- follow an original target group with remaining people until the latest follow-up
- compare a segment of the group, where contributions have been discontinued, with ones where contributions continue
- compare the original target group with the 'entire' group consisting of cases with completed, ongoing or discontinued interventions at the time of the follow-up
- compare all these conditions between different target groups
- follow living conditions and needs for change as well as the activity's general contributions and impacts in 'general target groups' between different years over longer time periods.

The classic evaluation dilemma, to be able to study and understand which contributions have which impacts, is not primarily resolved by randomized control group studies - but by more complex pattern analyses, whereby the patterns of interventions conducive to the patterns of target achievement in various spheres of life can be studied in selected target groups. (Sandell 1998. US Patent: Multidimensional Pattern Analysis, Analysity / Sandell 2001).

Information compilation through structured dialogic interviews with users

The content of evaluation is based on the practice that personnel, together with users through structured, dialogic interviews, examine and document the following:

Background information; Life situation and needs for change in different spheres of life; Own resources; Contributions by different Change-implementing actors; Changes in the actual life situation; Assessment of Change; Target achievement; and the significance of Change-implementing actors / contributions as well as information about insights into target groups.

The dialogic interviews follow a cross-structure with spheres of life at three main levels:

The Social situation (Structural level), including spheres of life such as employment, housing, finances, education, etc; Relations (Interpersonal level) with life spheres such as child-parent, family, couples, close persons etc; as well as Personal (Individual bio-psychological level) with life spheres such as physical/mental health, self-image, dependence etc.

As regards the spheres of life, the interview follows a general aim and time perspective:

Life situation at Start – gives a description of the situation concerning the clients' actual life situation at the time 'before' possible interventions began.

Target for desired change – expresses the need for change in different spheres of life in connection with the start situation and is given irrespective of whether the activity makes any contributions.

Follow-ups on actual life situations – examine the new/actual life situations at the time of follow-up. The follow-up situation with possible changes may be compared with the start situation, irrespective of whether the targets and/or contributions are achieved.

In joint dialogic reflections during follow-ups, the following items are assessed:

Experienced change – which is given irrespective of whether change is desired or not and is assessed on a scale of positive, none or negative.

Target achievement – if needs for change have been put forward in the form of 'targets for desired changes'; assessed on a scale of excellent, good, partly or none.

The significance of change-implementing actors for the actual changes accomplished by care / casework efforts in the users' lives concerning actors that have in some form been actively involved in relation to the clients' needs for change.

The structure of evaluation with its content, levels, spheres of life and time perspective constitutes a framework for conversations. This means that all interviews in all cases follow the same basic structure and cover the same main and subsidiary headings. At the same time information gathering may be carried out in an individually adjusted order and in free conversation under these headings, which in terms of the content correspond with, and in form resemble, the ordinary practice of welfare work in client contacts.

The structure aims at contributing to an overall view and insight, enabling an increased understanding of people's life situation and needs for change – which, in turn, augment the validity of evaluation. The selected structure, with spheres of life divided at various levels, makes it essentially applicable to all people in all conceivable target groups in all welfare organizations, irrespective of their individual situations, resources and needs for change. There are no individuals on earth to whom these headings could not be applied.

In the evaluation-dialogues between the professional welfare worker and individuals / families, one should strive for openness, respect, active listening and empathic treatment. The ambition of this type of conduct is to strengthening the users' resources, responsibilities and self-determination. This dialogue also gives personnel and users an opportunity to metacommunicate about the intention, form and content of the conversation for the mutual and reciprocal clarification of the meaning of what is being 'asked' and 'answered'. Thus the dialogue enables an increased understanding of the persons' life situation, needs for change, own resources and the actual changes. In this way, these dialogic interviews can be integrated into casework and support co-operative relations in effecting care and actions for change.

The external forms of the dialogic interviews – such as duration, place, tempo etc. – should naturally be adjusted to the clients' situation, needs, age and developmental stage. For example, it is possible to conduct information gathering on several occasions, take breaks in conversation or have a cup of coffee, meet at the family's home, go for a ride in a car or take a walk etc. Consequently, it is not necessary that everybody remains seated and unsmiling in order for the information collection to be serious. It is a mutual endeavour to understand the clients' actual life situation and needs for change and therefore the forms of this information gathering must be adapted accordingly in a flexible manner.

The type of information to be documented within each sphere of life (with or without a computer program) constitutes naturally a summary – void of the personal nuances and deeper perspectives that may be part of a 'complete' conversation. This state of affairs, i.e. that a deeply personal conversation for the sake of documentation must be reduced to a given value for a variable within a given sphere of life, may be perceived as 'superficial' by service users and professionals alike. This simplification is, however, necessary so that all information from many several cases can be subsequently aggregated to meaningful results for overview. Even if this deep information is left out from documentation, it has however, not been gathered 'without due cause' - as it remains present in treatment relations and (if so desired) logged in a diary – for both parties to consult in further efforts to effect changes.

Evaluation entails the compilation of various types of information.

In the interview (as in subsequent result analysis), one should be aware that in different spheres of life the type of data processed is qualitatively divergent.

The external social situation, such as employment, housing and finances, some personal health-related data and certain 'registered' conditions (for example sentences for offences committed), may be defined as 'objectifiable' facts. Here, the mutual task is to uncover truthful information about the actual conditions in the user's life. On the other hand, relations with the family and close persons as well as personal conditions, such as psychological health and self-perceptions, are primarily statements about the individual's / family's values and perceptions. Even if the dialogue contributes to a mutually increased understanding, the quality of relations should ultimately be decided on the client's perceptions.

In the target for desired change of needs analysis, these real needs should not be confused with 'desires for contributions' or market-oriented ideas, such as 'demand / satisfied customers'; instead they must concern the desired changes in the clients' life situation.

In mutual reflections on follow-up, concerning changes and target achievement, an effort is made to compare the changed life situation with the initial targets. These assessments of change and target achievement constitute a comprehensive analysis of the mutual dialogue – between the client's / family's perceptions and the professionals' knowledge and experiences. Therefore, in this integrated evaluation no information is collected

behind the user's back or done over his or her head but everything takes place openly and in mutual co-operation.

Experiences of what is required to make integrated evaluation possible

In general terms and irrespective of the area of activity and model, a serious evaluation should be:

- grounded in scientific theory in its aims, cognitive view, researcher role, context,
- perspective and cognitive interest as well as all-embracing validity and reliability
- well thought-out research methodology in relation to what is evaluated, in which context,
- how, when and by whom as well as by which qualitative and/or quantitative methods.
- `anchored in reality´ through 'tested experience', which means that both the evaluation model as such and the contributions evaluated must have been tested with clients and that the information obtained, whenever possible, can be implemented and utilised.

In addition, evaluation in welfare organizations and activities must take account of the aims and specific prerequisites of welfare and therefore must essentially:

- examine the client's life situation, own resources and needs for change,
- follow the life situation and needs for change over time to be able to examine changes in clients' lives over a given evaluation period,
- know which interventions have been made in different cases during this evaluation period to be able to relate the results to the contributions,
- be able to aggregate life situations, needs for change, own resources, contributions and results in isolated cases at the group level for drawing conclusions that can be generalised
- be able to use the aggregated information for analysing the benefits from activities for the client and be able to relate the results from changes to the aims of the activities,
- be able to give activities feedback on the information obtained for own quality development.

Integrated evaluation, which has emerged from many years of development in co-operation with many different welfare organizations and activities, has naturally met with stumbling blocks and challenges calling for both scientific consideration and field experiments.

In addition to the above-mentioned general and welfare-related evaluation requirements, experience has shown that the following complementary requirements must also be placed concerning the direction, content and implementation - so that evaluation is actually possible to integrate into ordinary welfare activities, such as the social work in an entire city.

The aim of evaluation must be congruous with the aim of welfare.

In statutory and democratically steered welfare systems, it is neither possible nor desirable that personnel should conduct an evaluation of their own work that would have aims other than those of the welfare organizations. Moreover, it is not possible to distinguish between actual client service and efforts to implement changes from information gathering done together with clients. These two aspects of 'one and the same' job must therefore have congruent aims and approaches.

The volition, competence and commitment of professionals are necessary for implementing integrated evaluation.

While integrated evaluation naturally must be grounded in and decided by policymakers and operational management, it cannot be 'ordered'. The model's fundamental approach, with empathic listening and co-operation with the clients in dialogic interviews, is crucial to the quality and credibility of the information obtained. Curiosity to know the results and desire to improve one's professional competence is as well needed so that the information yielded by evaluation can actually be used for developing the quality of activities. This is not possible without the commitment and involvement of the evaluating professional personnel .

Personnel must have time for testing and reflection and their possible doubts must be dealt with respectfully.

Professionals sometimes express concerns for taking part in evaluation because it takes too much time "from work", possibly "interferes" with relations with the users and perhaps could jeopardise the clients' integrity. In some cases, there are also doubts about one's own competence to conduct dialogic interviews with clients and, in exceptional cases, there may be reluctance to openly show the results of the activity – "just think if it shows that we're not doing anything worthwhile". In principle, these misgivings can only be addressed by taking them seriously and by giving the personnel an opportunity to test the model together with the clients – a practice that almost invariably dissipates these concerns.

The users must be informed and made aware of the aim of integrated evaluation so that their co-operation is 'genuine'.

User co-operation is usually no problem, as the clients, almost without exception, are positively inclined to co-operate in integrated evaluation, when it is made clear to them that evaluation aims at better activities for the added benefit of the user, that it takes place within ordinary work and in principle encompasses 'all clients', and that their integrity is totally protected when the results are presented. When they afterwards notice that the dialogic interviews are not about being questioned as an 'object' but a joint enterprise to seek important knowledge, the users most often think very positively about both co-operation, needs analysis, implementing changes and the subsequent follow-ups.

The forms and approaches of evaluation must support the users' resources and commitment to change and contribute to the development of salutogenic treatment relations.

The client's own resources can be reinforced, if needs analysis does not solely focus on problems and shortcomings but also emphasises resources in the individual, family and social network. With a sound outlook and information compilation in a dialogic form with open-minded listening and respect, it is possible to avoid an objectifying distance that may risk to undermine the care / treatment relationship.

Policymakers and operational management must lend active support to integrated evaluation by asking for the information gathered.

The relevant decision-makers must stand behind the aims of evaluation and be interested to learn "how the clients are doing" as well as be prepared so that this information can have an impact on the ongoing programme, plan and budgetary work. To decide on evaluation but lack the determination to utilise the knowledge obtained would of course result in insufficient credibility that in the long run would compromise personnel's commitment.

The activity should 'own' the information and results from evaluation and should itself decide which target groups are analysed and what are the conclusions to be drawn.

'Self-interest' is needed so that the leadership and personnel of a welfare system commit themselves to information compilation, result analyses and conclusions that lead to quality development. If personnel and leadership perceive evaluation as belonging to somebody else, with analysis separated from the activity and with no connection to how the work with one's clients could be improved, it is only understandable that commitment starts to erode.

Responsibility for evaluation should follow the division of responsibilities in the activity.

Elected officials and operational leadership should clearly decide to engage in integrated evaluation and that feedback on its results will be forthcoming in well thought-out forms. Responsibility for conducting the evaluation should then follow the regular order of making decisions, with supervisors responsible for implementing it and giving feedback on the results in their respective areas of responsibility. This means that evaluation belongs to the 'regular´ organization and activities, providing the necessary stability over time and reducing the risk that evaluation becomes too much 'person related' or viewed as a temporary 'project'.

Personnel must have enough time to conduct information compilation for evaluation within the timeframe of their ordinary work.

The concrete forms of data collection, documentation and aggregation must not take an excessively large proportion of working hours. This as such is an important reason for utilising data processing in evaluation, providing, for example, time-saving headings and preplanned lists. This also leads to an inherent problem, how to best combine two aspirations that do not always mix together. On the one hand, evaluation must be sufficiently broad, deep, accurate, etc. in order to generate serious and credible information, but on the other hand, it should not be too comprehensive and cumbersome to conduct within the framework of ordinary work. Therefore, it is not possible to choose the 'easy way' with 'the maximum data compilation'; instead, one should find the optimum, relevant, joint and adequate ´information base´- required for drawing conclusions whether the activity is worthwhile and fulfils its goal.

> *Personnel must bear with the simplification of information gathered in dialogue with clients in evaluative documentation.*

Simplification is a necessary result of compromises made in terms of scope and time and it becomes evident as personal dialogue is documented in some form, snd maybe even more by using computer programs with ready-made headings and lists. Nevertheless, frustration usually ease off when it is later understood that even if the many details and individual experiences captured in the dialogue and co-operation with the users had been documented, they would eventually be ´lost´ when the aggregated results are presented.

> *IT- support for documentation, analysis and presentation of the results must function in a user-friendly, time-saving and graphically clear manner.*

It is important that the computer programs used for evaluation run smoothly and that IT- support operates well in relation to the personnel's other tasks and record-keeping software. Otherwise there is a risk that computer troubles take precious working hours and hinder smooth co-operation with clients. Naturally, the documentation software, databases, analysis and presentation tools should come complete - with adequate and comprehensive security, integrity, functionality, capacity and IT support.

> *Feedback on the results and conclusions should be directed at the users as well as personnel, unit, operational leadership and relevant policymakers.*

Feedback to the clients is automatically incorporated in every follow-up and crucial for ethical and judicial reasons, but also because it contributes to an open co-operative relationship in client service. Feedback to the unit and personnel in question is necessary, partly in order to enable result analyses and jointly drawn conclusions, partly because the efforts specified in the evaluation must be perceived as serving long-term self-interest, meaning that staff members themselves will get something in return, in forms of enhanced competence and improved activities.

Integrated evaluation should be linked to programme and planning work done within the welfare organization and be included in the yearly reporting on activities.

The information gleaned by evaluation will hardly have any use for the quality development of organizations, if it leads a solitary existence parallel to the actual decision-making processes – generating interesting information that has absolutely no impact on activity planning and priorities, methodological development, further training, etc. Giving feedback on the information provided by evaluation to operational leadership and political committees must therefore take place in well thought-out forms suiting the activities and be 'time-tabled' to take place at least once a year so that information about the users is brought up at the same time when the programmes, plans and budget are being discussed and decided upon.

Integrated evaluation puts increased demands on validity and reliability.

The demands for the validity, relevance, reliability, authenticity and accuracy of the information gathered – that concern all serious evaluations – are even weightier in integrated evaluation. When this information compilation belongs to everyday welfare work, it is not enough that the information as such is relevant, adequate, plausible and correct for the type of knowledge asked. In addition, the clients must be able to understand this information and it must be implemented in the ongoing change-effecting work with the people, families and target groups as well as be conducive to the quality development of the welfare organization. This is another decisive reason for having dialogue as the primary form of information gathering, since it enables continuous clarification of the content and meaning to improve intelligibility and applicability.

To create a learning organisation

Integrated evaluation aims at long-term quality development through the knowledge obtained for the clients' benefit. Thus, evaluation ultimately has justification, if it leads to such improvements in activities that have positive consequences for the service users' lives. However, this ambition cannot be attained by an isolated evaluation done during a limited period of time, regarding a specific target group and aiming, for instance, to establish that a certain method is the 'best' solution to a certain problem.

'Reality' in society, welfare work and the clients' lives is subject to dynamic change. Information that seems to be valid for a given method in a certain target group and in some temporal context is liable to be partly obsolete already when presented, because in the meantime reality has already moved on. This dilemma, which essentially concerns all forms of evaluation, should be dealt with a humble insight that there is no simple ´truths´ or linear causality in the clients' lives or in multifaceted welfare activities. This should not, however, lead to resignation in the face of the complex reality

and to disbelief in evaluation information. As a professional welfare worker, one must appropriately seize the currently best information about the impacts of various interventions on different needs for change in the target groups. Even though this information may not be true forever, it is preferable to fumbling in the dark, having no evaluative information at all.

The problem with evaluation in a dynamically changeable reality can, however, be addressed by developing welfare activities into learning organisations. The fundamental idea behind the concept of a learning organisation is that both existence and information are regarded as a 'process' rather than as a 'state'. Compilation of information must also take place 'within the process' in order to obtain information based on experience, enabling continuous feedback to activities for yet new attempts, information and feedback. This reduces the risk of clinging to counterproductive old 'truths' or getting lured by fashionable trends irrelevant to evaluation knowledge. Learning of course embraces other dimensions as well, such as basic and further education, methodological and process guidance, job supervision, human resources development, studying the literature, research reports and the like.

Thus a genuine learning organisation can be established, if the evaluation continuously documents and reports on the information and experiences of the activities for new efforts to improve that will in turn be evaluated and so on. This enables a continuous information-flow that can follow the complex and changeable reality in which we together try to co-operate, change, reflect and learn 'within the process'.

This ambition also constitutes a decisive reason why evaluation should be integrated into the regular welfare organizations and activities and conducted in dialogue with the users.

Increased interest in integrated evaluation

The model of integrated evaluation with the advanced and relevant IT-support is currently being implemented in some of Swedish municipalities - in the social services and care for individuals and families through advice and support, consulting, financial matters, employment and misuse as well as inquiry, treatment and placement concerning children, young people and adults. Much interest has also been shown among other welfare organizations, such as family counselling and casework, family courts, elderly care, rehabilitation, immigrant services, child and juvenile psychiatry, welfare for schoolchildren etc. In addition, this practice has attracted interest in international contexts, such as the International Family Therapy Association (about 60 member countries) and the Systems in Transition (NGO status in the UN) as well as in the 'social welfare sector' in the USA.

Over the years, the model has been the object of some meta-studies. The original research programme BRA/Behovs och resultatsanalysis (Analysis of needs and resources), was conducted under the scientific leadership of Sandell (1988) and researched by Frankenberg, Sandell and Thylefors (1990), Karlsson and Pousette (1992) and Eriksson and Karlsson (1998). An ongoing comparative study by Alexandersson (2002 -) discusses the actual model that is operational in some municipalities. According to the recurrent

patterns emerging from these studies, the users adopt a positive attitude towards the dialogue and treatment within the model, while the leadership and personnel, sometimes after initial hesitation, having practised the model together with the users, also have formed a genuine positive opinion. This is also the common experience in all R&D projects on the model throughout two decades.

On the other hand, integrated evaluation has been met with silence or distrust in academia, even though no serious criticism has been voiced concerning its ideology, theory or method. This is probably due to the fact that many researchers lack the integrated model's 20 years of positive experience in direct co-operation with various welfare organizations and activities and confidence in professionals' competence to engage in dialogue with the users. In recent years though,, it has seemed that the academic tide is turning and that the model is increasingly often invited for presentation in university instruction, conferences, central authorities and the regional R&D context. Maybe there is also a rising insight, considering time, money, flexibility and advanced analyses, that IT-support is needed if it should ever be possible to evaluate the large, divergent, multilevel, ongoing, ordinary welfare organizations.

Finally, the paradigm advocated by integrated evaluation is not primarily about new methods, organisational forms or IT support. Basically this paradigm is about a fundamental view on welfare values, theory and praxis, where the personnel are regarded as professionally competent and the users as responsible fellowmen equipped with resources of their own.

The model is founded on many years' positive experiences - showing that the personnel, in a structured dialogic co-operation with the users, can act as compilers of important knowledge - in an evaluation that is integrated into ordinary welfare work.

References:

Alexandersson, K. (2002). Utvärderingens praktik. Dalarnas forskningsråd, Falun.

Anastas, J. W. (1999). Reseach Design for Social Work and the Human Services. New York: Colombia University Press

Anderson, H. Goolishian H. (1999). Från påverkan till medverkan. Stockholm: Mareld.

Andersen, T. (1999). Reflekterande processer. Stockholm: Mareld

Armelius & Armelius (1999). Kausalfrågor och effektfrågor vid utvärdering av behandling. Stockholm: SIS.

Boscolo L., Bertrandi (1998). Genom tidens lins. Stockholm: Mareld.

Eriksson B & Karlsson P-Å (1998): Utvärdera bättre – för kvalitetsutveckling i socialtjänsten. Gothenburg: Gothia förlag.

Frankenberg, K., Sandell, G., Thylefors, I. (1990). Metastudie om förändringsprocesser – ett FoU-projekt inom Bra- programmet. City of Gothenburg.

Karlsson, P-Å. (1992) Går det att bistå med BRA resultat ? Bra-programmet. City of Gothenburg.

Kvale, S. (1997). Den kvalitativa forskningsintervjun. Lund: Studentlitteratur.

Lundsbye, M., Sandell,G., Währborg,P., Ferm, R., Pettit,B., Fälth,T. (1985- 2000) Familjeterapins grunder – ett interaktionistiskt perspektiv. Stockholm: Natur & Kultur.

Morén, S. (1996). Att utvärdera socialt arbete. Stockholm: Publica.

Sandell, G. (1985). Psykosocialt förändringsarbete. Doktorsavhandling, Institutionen för socialt arbete, Göteborgs universitet. Bok på Liber förlag, Stockholm.

Sandell, G. (1985). FoU-policy för socialtjänsten. City of Gothenburg.

Sandell, G. (1986). Policy för en ny förvaltningskultur. Social Services, City of Gothenburg.

Sandell, G. (1988). BRA-programmet / Behovs- och resultatanalys inom Socialtjänsten. Programskrift samt beskrivning av fyra delprojekt: Bra-barnomsorg, Bra-hemtjänst, Bra-bistånd / IFO samt Metastudie av Bra-programmet. Social Services, City of Gothenburg.

Sandell, G., Berglund, M. (1993). Bedriver vi en effektiv verksamhet? Behov, resultat och kvalitet. Modell för integrerad utvärdering. Sv. Kommunförbundet. Kommentus.

Sandell, G., Feldt, V. et al. (1997). 110 kriminella pojkar. Vad blev bättre ?

Stockholm stad: Maria ungdom / Kriminalvårdsprojektet.

Sandell, G., Olsson, J. (1997). Bli sedd, välja väg och förändras. – Utvärdering av utredningsarbetet vid Ljungbacken, en ungdomsinstitution inom SiS. Stockholm:

Sandell, G. (1998) Göteborgarnas psykiska ohälsa. City of Gothenburg: Samhällsmedicinska expertrådet. Folkhälsorapport Nr 5 , 1998.

Sandell, G. (1998). Multidimensional Pattern Analysis. US 6.282.532.B1 / 2001. International Patent for an IT model for integrated evaluation and pattern analysis. Washington: United States Patent and Trademark Office.

Sandell, G. (1998 – 2003). Integrerad utvärdering. Ideologi, teori, metod och IT-stöd. Paradigm och modell för utvärdering och kvalitetsutveckling inom välfärdsverksam-heter. Bok och bildspel samt Info till dataprogram / Analysity

Tomm C. (1989). Systemisk intervjumetodik. Stockholm: Mareld.

Westlund, P. (1999). Lära av varandra : kollegiegranskning, kvalitetskartor, fokusgrupper. Fokus-rapport 1999. 7, Kommunförbundet Kalmar

White M,. Epston D. (2000). Narrativ terapi. Stockholm: Mareld.

Övereide H. (1998). Samtal med barn. Oslo: Höjskolförlaget.