

JMDE

Journal of MultiDisciplinary Evaluation

Number 2, February 2005

Part I

Editors

E. Jane Davidson & Michael Scriven

Associate Editors

Chris L. S. Coryn & Daniela C. Schröter

Assistant Editors

Thomaz Chianca

Nadini Persaud

John S. Risley

Regina Switalski Schinker

Lori Wingate

Brandon W. Youker

Webmaster

Dale Farland

Mission

—The news and thinking
of the profession and discipline of evaluation
in the world, for the world—

A peer-reviewed journal published in association with

The Interdisciplinary Doctoral Program in Evaluation
The Evaluation Center, Western Michigan University

Editorial Board

| | |
|--------------------|--------------------------|
| Katrina Bledsoe | Shawn Kana'iaupuni |
| Nicole Bowman | Ana Carolina Letichevsky |
| Robert Brinkerhoff | Mel Mark |
| Tina Christie | Masafumi Nagao |
| J. Bradley Cousins | Michael Quinn Patton |
| Lois-Ellen Datta | Patricia Rogers |
| Stewart Donaldson | Nick Smith |
| Gene Glass | Robert Stake |
| Richard Hake | James Stronge |
| John Hattie | Dan Stufflebeam |
| Rodney Hopson | Helen Timperley |
| Iraj Imam | Bob Williams |

Table of Contents

PART I

Editorial

| | |
|---|---|
| In this Issue: JMDE(2) | 1 |
| Marketing Evaluation as a Profession and a Discipline | 3 |
| <i>E. J. Davidson</i> | |

Articles

| | |
|--|----|
| Monitoring and Evaluation for Cost-Effectiveness in Development Management..... | 11 |
| <i>Paul Clements</i> | |
| Network Evaluation as a Complex Learning Process | 39 |
| <i>Susanne Weber</i> | |

Practical Ethics for Program Evaluation

| | |
|---|----|
| Client Impropriety..... | 72 |
| <i>Chris L. S. Coryn, Daniela C. Schröter, & Pamela A. Zeller</i> | |

Ideas to Consider

| | |
|--|-----|
| Managing Extreme Evaluation Anxiety Though Nonverbal Communication ... | 76 |
| <i>Regina Switalski Schinker</i> | |
| Is Cost Analysis Underutilized in Decision Making? | 81 |
| <i>Nadini Persaud</i> | |
| Is E-Learning Up to the Mark?..... | 83 |
| <i>Oliver Haas</i> | |
| The Problem of Free Will in Program Evaluation..... | 102 |
| <i>Michael Scriven</i> | |

In this Issue: JMDE(2)

Michael Scriven

The journal homepage has had over 6,000 hits, and there have been around 2,500 downloads of all or part of the first issue. Our list of 932 people who want to be notified of new issues now includes residents from more than 100 countries. The current issue is a bit longer: it runs over 170pp. but you can download just the parts that interest you. Here are some highlights.

- There is an editorial by Jane Davidson on the perception of evaluation by others, and what we can and should do about it.
- One of the major articles is by Paul Clements, who raises serious concerns about the crucial matter of how the big (U.S. and other) agencies are evaluating their vast expenditures on development programs overseas. He's unlike most critics in two respects: (i) he went to Africa to check things out on the ground for himself, and (ii) he suggests a way to raise the standards considerably. You will no doubt realize that both the problem he writes about, and his proposed solution, have obvious generalizations to other areas of public and private investment.
- The other major article is from Germany, in which Susanne Weber sets out an approach to monitoring and evaluation based on current abstract sociological theorizing. Her approach also bears on systems theory and organization learning, in case those are interests of yours. That article and the other German contribution (on the evaluation of online education) are interesting not only for

their content, but for the sense they provide of how evaluation is seen by scholars in Europe.

- We introduce a new feature—“Ideas to Consider”—for short pieces, selected by the editors and ideally just of memo length, that canvas ideas we think deserve attention by evaluators. There’s a quartet of these to kick the feature off: one on the still-persisting shortage of cost analysis in published articles and reports on evaluation, one on the role of body language in creating and countering evaluation anxiety, one on approaches to evaluating online education, and one on the tricky problem of how to evaluate programs (or drugs) which depend on the motivation of users for their success (should attrition rate count as program failure or subject failure?).
- Our strong interest in international and cross-cultural evaluation continues with an update on several of our previous articles covering evaluations in regions and publications around the world. The review of evaluation in Latin America and the Caribbean in the last issue has already been reprinted in translation, as it well deserved, and its sequel here tells an impressive story of activity in that region. The sixteen articles in this section tell a remarkable story: evaluation is changing the world and the world is changing evaluation!

- MS

Marketing Evaluation as a Profession and a Discipline

E. Jane Davidson

[Davidson Consulting Limited](#), Aotearoa/New Zealand

It can be a bit like pushing sand uphill with a pointy stick, as they say here in New Zealand. One of the great challenges in developing evaluation as a discipline is getting it recognised as being distinct from the various other disciplines to which it applies. In this piece, I offer a few reflections on the challenges with this, recount a story where a group of practitioners from outside the discipline actually sat up and took notice, and propose some possible solutions for moving us forward.

It's all very well for us to come together in our evaluation communities around the world and talk to each other about our unique profession. Not that there isn't a lot to talk about. After all, we are still working on building a shared understanding of what it is exactly that makes evaluation distinct from related activities such as applied research and organisational development. But with a little more application, I am hopeful we can persuade enough of a critical mass to call it a reasonable consensus.

Meanwhile, it seems to me that a more difficult yet equally important task is to articulate clearly to the outside world—to clients and to other disciplines—what it is that makes evaluation unique.

Right across the social sciences and in many other disciplines where evaluation is relevant in more than just its intradisciplinary application, it seems that the vast majority of practitioners consider it to be part of their own toolkit already, albeit

often under a different name. Most of these practitioners consider evaluators delusional when we suggest that evaluation is sufficiently distinct to call a profession, let alone an autonomous discipline.

Here's a fairly typical response, in this case from an industrial/organisational psychologist:

[A] discipline evaluation is not. Disciplines are systematic, coherent, founded more often than not on sound theory, and offered as programs in accredited colleges, universities, and professional schools. Evaluation, without detracting in the least from its multitude of contributions and creative authors and practitioners, is not systematic, coherent, theory-driven, and offered—oh perhaps with an exception here and there—as a program of study at institutions of higher learning. Evaluation is a helter-skelter mishmash, a stew of hit-or-miss procedures, notwithstanding the fact that it is a stew that has produced useful studies and results in a variety of fields, including education, mental health, and community development enterprises.¹

Industrial and organisational psychology is a relatively young discipline itself, but obviously not quite young enough for its practitioners to recall the struggles they must have had in the late 1800s and early 1900s. Industrial psychology, which focuses primarily on personnel selection and ergonomics/human factors, grew out of a blend of industrial engineering and experimental psychology. The first doctoral degrees in industrial psychology did not emerge until about the 1920s.

¹ For the full article, see R. Perloff's (1993) [“A potpourri of cursory thoughts on evaluation.”](#)

It seems likely to me that the fledgling discipline of industrial psychology had its share of critics in those days. Perhaps it was even called a “helter-skelter mishmash.” There was probably a lot of dissent in the ranks as to whether it really was any different from industrial engineering, measurement, experimental psychology, and a host of other disciplines. And I am sure there were furious debates about the definition of industrial psychology itself.

Was it (or would it have been) reasonable to declare industrial psychology a discipline even though most insiders didn’t agree on its definition, underlying logic, or the soundness of its theories? How much shared understanding constitutes a critical mass?

There’s something of a “chicken and egg” argument here. It seems to me that little progress in theory or practice can be made beyond a certain point without first declaring evaluation to be a discipline and then seeing what develops. Sure, not everyone will buy the idea initially, but there’s no point being put off by those who like to throw their hands in the air and declare the whole exercise impossible. These things take time, open minds, thinking and rethinking.

Whether or not we have the courage and conviction to declare ourselves a discipline at this point, I think it’s fair to say we have a critical mass who are quite clear that evaluation is at least a professional practice with a unique skill set that is honed with reflective practice and other forms of learning. The challenge here is convincing non-evaluators (such as the I/O psychologist quoted earlier) of this.

Consultants are a particularly hard nut to crack. Often trained to the graduate level in business and/or the social sciences, the almost universally held perception is that all one needs to do evaluation is some content expertise and perhaps a few measurement skills (and accounting skills).

What could possibly make seasoned professionals such as management consultants sit up and take notice of evaluation?

Let me set the scene. A client organisation had put out an RFP asking for an independent evaluation of a leadership initiative. Interestingly, the RFP specifically stated that the client was looking for an evaluation expert with content expertise rather a content expert (e.g., a management consulting or industrial/organisational psychologist) with evaluation experience. This is very unusual in the evaluation of leadership initiatives. Most clients are unaware that there is such a thing as evaluation expertise, as distinct from the applied research skills a well-qualified management consultant or organisational psychologist might possess.

Of 22 initial expressions of interest in the contract, just two (yes, 2!) of these were from people who identified as evaluators and participated actively as members of the [national and international] evaluation community. This was despite unusual efforts on the part of the client to attract expressions of interest from evaluators. Rather than simply posting the RFP on the usual electronic bulletin boards, which they had heard good evaluators do not usually respond to, they also sent out direct emails to evaluators who had been recommended by other evaluators and had the notice posted on an evaluation listserv.

[It is interesting to note that the process used by the client to specifically target evaluators closely mirrors best practice for the recruitment of top-notch job candidates, especially underrepresented groups—don't just use the regular channels that yield the same old candidate pools; go to where you know the right people are and personally encourage them to apply.]

The client in this case, under the guidance of an evaluator not bidding on the job, used a creative and unusual process to select the contractor. Rather than asking

shortlisted bidders to submit the usual 20-page proposal, the selection team invited them to a face-to-face meeting where they could present their thoughts on the evaluation. This was because credibility was a key element of the evaluation, which the client felt couldn't accurately be gauged without meeting the evaluator face to face.

An added benefit of the face-to-face interview approach was that it increased the odds of both attracting and identifying a "real" evaluator. In a small community such as New Zealand, the vast majority of evaluators are solo practitioners who often partner with others for particular pieces of work. As such, they have inadequate resources to devote to compiling lengthy, slickly presented proposals that have less than an even chance of being successful. In contrast, larger consulting firms who do *not* have evaluation as their primary function are far more likely to have an extensive library of proposal templates and a number of junior staff trained in writing proposals. Therefore, the standard written proposal solicitation process is far more likely to yield bids from content experts than from evaluators.

Prospective contractors were asked to submit a number of supporting documents for the interview, including an outline of their "quality assurance procedures." The proposed quality assurance procedures turned out to be one of the more telling pieces of information. After all, what better way to understand an evaluator's grasp of his or her profession than to ask how his or her work should itself be evaluated? One case in point was a large, multinational business consulting firm ("Firm X") whose quality assurance procedure consisted of appointing one of their independent auditors to oversee the evaluation.

In the final round, only the two “actual” evaluators passed the interview process and made it onto the final short-shortlist for being awarded the contract. When the final decision was made, the runner-up was told the background and qualifications of the successful bidder—and immediately recognised who the competitor was (New Zealand being a small evaluation community). By chance, the two met up a few days later and had a chuckle when they finally connected the dots.

In contrast, Firm X was, by all accounts, extremely surprised not to make even the final short-shortlist of two. To their credit, they did send a junior employee to see the client to get feedback about why their bid had been unsuccessful. They were even more surprised to be told that the main reason was because they were *not evaluators*. And no, “audits” and “reviews” of the type they were well versed in were not the same as high-quality evaluations. The consultants from Firm X were flummoxed!

Firm X asked who had been awarded the contract to evaluate the leadership initiative, and were told. They then asked who the runner-up was. The client quietly pointed out that, if they really were evaluators (as they claimed to be), they would already have found that out through their extensive evaluation networks—in the same way as the top two contenders had found out about each other.

There is a wonderful lesson here for evaluation as it strives for recognition as a distinct profession and as a discipline. I think we’ve all tried convincing the colleagues in our content disciplines that what we do is unique, complex, more than just measuring a couple of variables of interest, and something worth paying attention to. And every now and then we get a breakthrough with our evaluation evangelising. But the reality is that evaluation-savvy clients will likely sell us more

converts among this audience than we could possibly manage for ourselves. There is nothing quite like being denied a contract for not being an actual evaluator!

What are some of the strategies we can use to educate clients? The simplest one that comes to mind is to highlight in our work what it is we are doing that is unique to evaluation. This might be serious and systematic attention to utilisation issues, the application of evaluation-specific methodologies not known to our non-evaluator colleagues, or the use of frameworks and models that have been developed specifically for evaluation. Whatever it is, we should be sure to highlight it in a way that makes it easy for a client to tell a “real” evaluation from the rest.

A second client education strategy is to seek opportunities to help with the development of evaluation RFPs. This was the case in the organisation I described, and it made a very substantial difference to how well the task was outlined, the selection criteria, the quality of the selection process, and client satisfaction with the outcome. Although the organisation was constrained by regulations about how an RFP process could be managed, good evaluative thinking allowed individuals within the organisation to generate a creative solution that led to the right result.

The third strategy for spreading the word about evaluation would be to follow the example of the Society for Industrial and Organizational Psychology (SIOP) in the States. Like us, I/O psychologists also have trouble getting the general public (especially managers in organisations) to understand what it is they are particularly skilled to do. In response to this need, SIOP has developed an extremely simple and straightforward leaflet, which it sends to members for distribution to managers they know. The goal was to have each member distribute the leaflet to five

managers. A copy of the leaflet may be viewed online at <http://siop.org/visibilitybrochure/siopbrochure.htm>

It is likely that by directing our educational efforts outwards toward clients, we will have the side effect of creating some better clarity within the evaluation profession, which will in turn let us make better sense to the outside world.

Monitoring and Evaluation for Cost-Effectiveness in Development Management

Paul Clements²

1. Development Assistance Requires a High Analytic Standard

In the Malawi Infrastructure Project, the World Bank planned to rehabilitate 1500 rural boreholes at a cost of \$4.4 million, with an estimated economic rate of return of 20%. At the project's Midterm Review, two years later, the rate of return was reduced to 14%, but the reasons for the reduction were not clear. The plan had anticipated that 85% of project benefits would come from the value of time the villagers saved that they would have spent collecting water, and 10% from the incremental water consumed. The Midterm Review estimated 31% of benefits from time savings, however, and 56% from incremental water consumed.³ No reason was given for reducing the estimate for time savings or for increasing the value for water consumption.

The World Bank's Fourth Population Project in Kenya aimed to decrease Kenya's total fertility rate to six births per woman by improving family planning services. The project was approved in 1990, and Kenya's total fertility rate fell from 6.4 in

² Corresponding author: Paul Clements, Department of Political Science, Western Michigan University, Kalamazoo, MI 49006, e-mail: clements@wmich.edu.

³ Carlos Alvarez, Ebenezer Aikins-Afful, Peter Pohland and Ashok Chakravarti, 1992, "Malawi Infrastructure Project: Mid-Term Review Report, September 1992," Lilongwe, Malawi: The World Bank, Appendix B. The economic analysis from the project plan comes from the Malawi Infrastructure Project's Staff Appraisal Report. I was given it upon agreeing not to reference it.

1989 to 5.4 in 1993. The project's Implementation Summary Reports consistently indicated that "All development objectives are expected to be substantially achieved," and a 1995 supervision report asserted that "The project development objectives have been fully met."⁴ Project activities, however, mainly supporting the National Council for Population and Development, were largely unsuccessful, and in 1994 a large part of the project budget was reallocated to the fight against AIDS. There were many other development agencies with family planning projects in Kenya, some with much stronger performance. Documents for the Fourth Population Project do not explain how its development objectives were related to the activities it funded.

The World Bank's Water Supply and Sanitation Rehabilitation Project in Uganda aimed to rehabilitate the water and sewerage system in Kampala, the capital city, and in six other major towns. Its plan calculated a 20% economic rate of return based on incremental annual water sales of \$5.5 million from 1988 to 2014. The completion report estimated actual returns at 18% because water production in 1991 was 10-20% below expectations.⁵ The project had indeed achieved its construction goals, but its efforts to strengthen the National Water and Sewerage Corporation (NWSC) had been undermined by the government's failure to raise water rates amidst hyperinflation and late payments on its water bills. The NWSC would have been unable to maintain the system without ongoing support, and

⁴ World Bank, 1995, "Form 590," (unpublished project implementation summary for Third and Fourth Kenya Population Projects), Washington, DC: The World Bank.

⁵ World Bank, 1991, "Project Completion Report: Uganda Water Supply and Sanitation Rehabilitation Project (credit 1510-UG)," Washington, DC: The World Bank, p. 30.

indeed by 1993, even with a major new project supporting the water company, it was once more operating in the red.⁶

These examples come from a blind selection of four World Bank projects that I studied for my doctoral dissertation.⁷ What is remarkable about these inconsistencies—an economic analysis in a midterm review that does not follow from the one in the project plan, development objectives that do not reflect project activities, an economic rate of return that anticipates 23 additional years of water sales based only on the current state of the infrastructure—is that even though at least the second two are at face value analytically incorrect, they are presented as routine reporting information, with no attempt to hide them such as in obfuscating language. Indeed they reflect common analytic practice in the international development community, and this common practice reflects a structural problem of accountability.

I would like to argue that the tasks undertaken by the large multilateral and bilateral donor agencies require a particularly high analytic standard, but several incentives that influence development practice—political incentives for donor and recipient governments, organizational incentives for development agencies, and personal incentives for managers—have led to positive bias and analytic compromise. These incentives are “structural” in that they result from the pattern

⁶ Paul Clements, 1996, *Development as if Impact Mattered: A Comparative Organizational Analysis of USAID, the World Bank and CARE based on case studies of projects in Africa*, doctoral dissertation for the Woodrow Wilson School of Public and International Affairs, Princeton University, p. 325.

⁷ Along with four projects of the US Agency for International Development and four from CARE International, all located in Uganda, Kenya and Malawi. The projects were selected based on descriptions of less than a page with no information on results.

of the flow of resources inherent in development assistance. The problem therefore requires a structural solution, and this paper proposes a possible solution involving a dramatic improvement in the quality and consistency of project evaluations. We can be confident that such an improvement is possible, first, because the evaluation problem facing development agencies has determinate features with specific analytic implications, and second, because a similar structural problem has already been addressed in the management of public corporations.

Sooner or later, development assistance comes down to designing investments and managing projects. Unlike private sector investments, development projects aim not to make a profit, but to improve conditions for a beneficiary population—to reduce poverty, or to contribute to economic growth. There is no automatic feedback such as in sales figures, and no profit incentive to keep managers on task. Typically one needs to strengthen existing institutions or to build new ones, and/or to encourage beneficiaries to adopt new behaviors and to take on new challenges. Yet in the project environment there is likely to be weaker infrastructure, a less well-educated population, and more risk and uncertainty than in the environments facing most for-profit enterprises. Furthermore, in places that need development assistance one cannot assume that institutional partners will be competent and mission-oriented. These conditions in combination place particular demands on development managers. Project managers need to maintain a unified conception of the project, its unfolding activities, and its relations with its various stakeholders, a conception grounded in a view of its likely impacts. Donor agency officials need a conception of the relative merits of many actual and potential projects, and an analysis that turns problems on the horizon for developing countries into programmatic opportunities.

The central challenge in the management of development assistance is to maintain this kind of consciousness—this analytic perspective—among the corps of professional staff. Some might like to think that development can be achieved by getting governments to liberalize markets or by getting local participation in project management, and these may well be important tactics. Intuition suggests and experience teaches, however, that there can be no formula for successful development. Each investment presents a unique design and management challenge. There are two problems in maintaining the will and the capacity to address this challenge: an incentive problem and one we can call intellectual or cognitive. The key to solving both problems, or so I will argue, is strong evaluation.

2. But Accountability in Development Assistance is Weak

2.1 Donor agencies are responsible for the success of their projects

According to the World Bank's procurement guidelines, "The responsibility for the execution of the project, and therefore for the award and administration of contracts under the project, rests with the Borrower."⁸ One might think that a development loan to a government is like a business loan to an entrepreneur. The donor agency makes the loan, but it is entirely the responsibility of the borrower government to spend the money. Whether a government manages its projects well or poorly, one might imagine, is primarily its own affair, with the donor providing technical assistance upon request. We know, of course, that this image is incorrect—donor agencies typically have the predominant influence over project design, and substantial influence over project administration—but it is useful to

⁸ The World Bank, 1985, "Guidelines: Procurement under IBRD Loans and IDA Credits," Washington, DC: The World Bank, pp. 5-6.

recall why this is so. One reason is parallel to a private bank's prudential interest in the management of its loans. As the World Bank's Articles of Agreement state,

The Bank shall make arrangements to ensure that the proceeds of any loan are used only for the purposes for which the loan was granted, with due attention to considerations of economy and efficiency and without regard to political or other non-economic influences or considerations.⁹

The Bank wants to be repaid, and it also has an interest in promoting economic growth and enhancing well-being in borrower countries, so it may take pains to see that its loans are well spent. Many loans are to governments with limited bureaucratic capacity in countries with inconsistent management standards, so the Bank must retain enough control to ensure that the projects it supports are properly administered.

By this logic we would expect relationships with bureaucratically stronger governments to be closer to the private sector model, and indeed some governments with coherent industrial strategies (consider South Korea in the 1970s) have succeeded in using Bank loans very much for their own purposes.¹⁰ Many development loans, however, are for projects at the edge of the borrower's

⁹ International Bank for Reconstruction and Development, 1991, "International Bank for Reconstruction and Development: Articles of Agreement (As amended effective February 16, 1989)," Washington, DC: The World Bank, p. 7.

¹⁰ See e.g. Mahn-Je Kim, 1997, "The Republic of Korea's Successful Economic Development and the World Bank," in Devesh Kapur, John P. Lewis, and Richard Webb, ed., *The World Bank: Its First Half Century*, Volume Two, Washington, DC: Brookings Institution Press, pp. 17-48.

frontier of technological competence, and the Bank (like other donor agencies) is a repository of expertise in the sectors it supports. The Bank also has demanding requirements for project proposals, and many governments have been unable independently to prepare proposals that the Bank could accept, particularly in earlier years when patterns of Bank-borrower relations were established. Therefore the Bank has generally taken primary responsibility for designing the projects it funds,¹¹ and the responsibility that comes with authorship cannot be lightly abandoned during implementation.

A second reason that donor agencies take an interest in how their funds are spent is that donor funds come from (or are guaranteed by) governments, and, the Bank's Articles of Agreement notwithstanding, governments do not release funds without taking an interest in their disposition. On one hand this political logic reinforces the prudential logic discussed above. Donor governments want their funds to contribute to the borrower's development, so they insist that donor agencies take responsibility for project results. Foreign aid is also, on the other hand, enmeshed in donor governments' general promotion of their foreign policy agendas.¹² It matters that the United States is not indifferent as to whether and when the World Bank will make loans to Cuba, and World Bank loans to Côte d'Ivoire have been subject to particular influence from France, the country's former colonial master.¹³ Bilateral aid is even more closely linked to donor government interests than aid

¹¹ Warren C. Baum and Stokes M. Tolbert, 1985, *Investing in Development: Lessons of World Bank Experience*, New York: Oxford University Press for the World Bank, p. 353.

¹² Paul Clements, 1999, "Informational Standards in Development Agency Management," *World Development* 27:8, 1359-1381, p. 1360.

¹³ Jacques Pégatiéan and Bakary Ouayogode, 1997, "The World Bank and Côte D'Ivoire," in Kapur, Lewis and Webb, ed., pp. 109-160.

through multilateral institutions. Not only from the donor side but from that of recipient governments too, the parameters of development spending cannot be understood merely in terms of the requirements for maximizing development impacts.

As intermediaries between donor and recipient governments, donor agencies are required to take more responsibility than private banks for managing the loans they make. The analogy with the private sector breaks down even further, however, when we consider the incentives governing a donor agency's management of its portfolio. The main cause for different incentive structures between donor agencies and private banks, of course, arises from differential exposure to financial risk. With private loans, the borrower and often the lender suffer a financial loss if the investment fails. With most development projects, by contrast, neither the donor nor the implementing agency faces a financial risk if impacts are disappointing. For projects funded by loans it is the borrower government, typically the treasury, that is responsible for payments. But the treasury seldom has control over individual development projects.

2.2 The usual watchdogs are not there to hold donor agencies accountable

The structural conditions of development assistance, therefore, create an accountability problem. Donor agencies have control over development monies but they face no financial liability for poor results (and no financial gain when impacts are strong). In this context their orientation to their task will depend largely on the demands and constraints routinely placed on them by other agents in their organizational environment, on the individual and corporate interests of their leaders and employees, and on the mechanisms of accountability that are institutionally (“artificially”) established.

In regard to external agents, Wenar notes that there has been a “historical deficiency in external accountability” for donor agencies.

Aid organizations have evolved to a great extent unchecked by the four major checking mechanisms on bureaucratic organizations. These four mechanisms are democratic politics, regulatory oversight, press scrutiny, and academic review.¹⁴

The electorates in donor countries want to believe that aid is helping poor people, but democratic politics also leads to pressures on donor agencies to support the agendas of well-organized interest groups.¹⁵ Some promote humanitarian and progressive agendas, but others have aims that create tensions with development goals. Generally, since the intended beneficiaries of aid cannot vote in donor country elections, the reliability of democratic politics as a source of accountability is limited. There has been significant regulatory oversight aiming to ensure that aid funds are not fraudulently spent, but external oversight of project effectiveness faces major practical hurdles. Aid projects are so widely dispersed, and the period between when monies are spent and when their results transpire is typically so substantial that effective oversight would require major bureaucratic capacity. Responsibility for project evaluation, however, has normally rested with the donor agencies themselves. This clearly leads to conflicts of interest, and it is the aim of this paper to suggest how these conflicts could be, if not removed, at least substantially ameliorated. Donor agencies have not, in any case, been subject to

¹⁴ Leif Wenar, 2003, “What we owe to distant others,” *Politics, Philosophy & Economics*, 2:3, 283-304, p. 296.

¹⁵ For example, American farmers have influenced U.S. food aid programs, which are overseen by the U.S. Department of Agriculture.

significant external accountability by way of regulatory oversight. Press scrutiny and particularly academic review, in contrast, have been significant sources of accountability, and academic studies have contributed to many foreign aid reforms. Given the strength of the political and bureaucratic interests that drive the programming of aid, however, and the above-noted dispersal of aid projects, scholars and journalists can only be expected to hold aid agencies accountable in a limited and inconsistent manner. Also, they are largely dependent, for information on aid operations on the donor agencies themselves.

Few who have spent much time with development agency personnel can doubt their generally admirable commitment to development goals, and the reforms this paper will propose depend heavily on the personnel's sustained interest in professionalism and effectiveness. Their behavior is also influenced, however, by their individual and corporate interests, and these interests take shape in the specific task environments that they face in their home offices and in the field. There are two aspects of the way their interests come to be constructed that are particularly relevant to the problem of accountability. First and most obviously, while institutional norms require donor agencies to maintain the appearance of a coherent system of responsibility for results, their institutional relationships require them to maintain the appearance that their operations are generally successful. They must evaluate, but it serves their individual and corporate interests if evaluation results are generally positive (or at least not often terribly negative). Since donor agencies have generally controlled their own evaluation systems, they have had the opportunity to design these systems in such a way that they would tend to reflect positively on the agencies themselves. Second, due in part to the long time span between the commitment of funds and the evaluation of results, internal personnel evaluations have tended to focus on variables only loosely

correlated with good results, and sometimes on variables that conflict with good practice.

2.3 Lacking secure accountability for results, other less relevant criteria inform resource allocation decisions

We will later consider some of the approaches donor agencies have taken to evaluation below. For the purposes of understanding the accountability problem in development assistance, it is enough for now to note that donor agencies have controlled their own evaluation systems. In the context of the general deficiency in external accountability, the priorities that have been enforced within donor agencies take on particular significance. Perhaps the most longstanding and sustained critique of donor agencies' internal operations involves the imperative to "move money."

The classic account of the "money-moving syndrome" is Tendler's *Inside Foreign Aid*.¹⁶ Focusing on the U.S. Agency for International Development (USAID) and the World Bank, Tendler identifies a "pressure to commit resources that is exerted on a donor organization from within and without," and finds that "standards of individual employee performance ... place high priority on the ability to move money."¹⁷ In the context of her organizational analysis, she gives several examples of aid officials knowingly supporting weak projects in order to reach spending targets.¹⁸ Tendler also finds, reinforcing the present argument about evaluation,

¹⁶ Judith Tendler, 1975, *Inside Foreign Aid*, Baltimore: Johns Hopkins University Press.

¹⁷ *Ibid.*, p. 88.

¹⁸ *Ibid.*, p. 88-96.

that in a political environment often hostile to foreign assistance, aid officials learned to self-censor reports that could provide ammunition for critics.

For writing what he considered a straightforward description of a problem or a balanced evaluation of a project, an AID technician might be remonstrated with, “What would Congress or the GAO [General Accounting Office] say if they got hold of that!?” ... Words were toned down, thoughts were twisted, and arguments were left out, all in order to alleviate the uncomfortable feeling of responsibility for possible betrayal. ... Such a situation must have resulted in a certain atrophy of the capacity for written communication – and, inevitably, for all communication through language.¹⁹

The World Bank typically required economic analysis of proposed projects, but Tendler found that many ostensibly economic projects were selected by non-economic criteria.²⁰ Much of the economic analysis that was carried out amounted to a “*post hoc* rationalization of decisions already taken.”²¹

While Tendler offers several political and organizational reasons to explain the money-moving imperative, I would like to emphasize what is absent from the organizational culture she describes. We do not find a sustained effort to consider how development funds can be employed to maximize their contribution to development. In such an environment we might expect well-intentioned professionals, once they win some organizational power, to act like policy

¹⁹ Ibid., p. 51.

²⁰ Ibid., p. 93.

²¹ Ibid., p. 95.

entrepreneurs, promoting their individual conception of a good development agenda in large measure despite the prevailing incentives. We might expect segments of a donor agency that have strong external allies to develop coherent agendas that they can implement themselves, as I believe reproductive health professionals at USAID have done. What we cannot expect, however, is that organizational decisions will routinely be taken on the basis of expected impacts.

The World Bank's "project approval culture" was recognized in its internal 1992 study, "Effective Implementation: Key to Development Impact" (popularly called the Wapenhans Report). The report cites a "pervasive preoccupation with new lending,"²² in part because "signals from senior management are consistently seen by staff to focus on lending targets rather than results on the ground,"²³ noting also that "[t]he methodology for project performance rating is deficient; it lacks objective criteria and transparency."²⁴ Although the report describes the Bank's evaluation system as "independent and robust," it finds that "[l]ittle is done to ascertain the actual flow of benefits or to evaluate the sustainability of projects during their operational phase."²⁵

Since the appearance of the Wapenhans Report, the Bank has moved increasingly to spending modalities that further dilute accountability for results. The two kinds of programs that have become most central to Bank strategies particularly in lower income countries are adjustment loans of various kinds (structural, sectoral) and

²² Portfolio Management Task Force, 1992, "Effective Implementation: Key to Development Impact," Washington, DC: The World Bank, p. iii.

²³ Ibid., p. 23.

²⁴ Ibid., p. iv.

²⁵ Ibid.

Poverty Reduction Strategy Papers (PRSPs).²⁶ Adjustment loans require borrowers to adopt free market reforms in order to better align economic incentives with development goals. They tend to operate on a wider scale than traditional projects, with more diffuse impacts. There is often a feeling that they are imposed, as the government receives the loan for policy changes it presumably otherwise would not have made, and they are often implemented only partially and inconsistently. These factors make them harder to evaluate. Poverty Reduction Strategy Papers typically push a larger part of the responsibility for evaluation onto the borrower government, and it seems that their more participatory approach to policy formation and implementation is intended to substitute, to some extent, for rigorous agency evaluation. They ask the government, as part of the process of generating a poverty reduction strategy, to identify a set of indicators for measuring the strategy's impacts. If the World Bank has had such a hard time ascertaining the level and sustainability of impacts from its own portfolio, however, it is questionable whether governments of low-income countries will be able to do much better.

3. Independent and Consistent Evaluation Can Improve Accountability and Learning in Development Assistance

3.1 The basic idea of the proposed evaluation approach

The problems discussed above present formidable obstacles to maintaining accountability in foreign assistance on the basis of program and project results. We should recall, however, what is at stake. In the absence of meaningful accountability there is little to counter-balance the pressures for aid resources to

²⁶ David Craig and Doug Porter, 2003, "Poverty Reduction Strategy Papers: A New Convergence," *World Development* 31:1, 53-69.

support political interests of donor and recipient governments, organizational interests of donor and implementing agencies, and personal interests of management stakeholders. The inconsistency and mixed reliability of evaluations have also undermined learning from experience, so the aid community has been slower than it would otherwise have been to identify successful strategies and to modify or abandon weak ones.²⁷

One way to address the historical deficit of external accountability, for example to push the focus of management attention forward from moving money to achieving results, and to improve the incentive and the capacity to manage for impacts, is to

²⁷ Indeed despite development agencies consistently reporting positive results from their overall operations, there have been persistent doubts about the basic effectiveness of development assistance at improving economic and/or social conditions in recipient countries. In their comprehensive 1994 review of foreign aid on the basis of donor agency documents, *Does Aid Work? Report to an Intergovernmental Task Force*, second edition, Oxford, UK: Oxford University Press, Robert Cassen and associates find that most project achieve most of their objectives and/or achieve respectable economic rates of return. A series of cross-country econometric studies, however, have failed to find evidence of positive impacts from foreign aid. These include Paul Mosley, John Hudson, and Sara Horrell, 1987, "Aid, the public sector and the market in less developed countries," *The Economic Journal*, 97:387, 616-641; P. Boone, 1996, "Politics and the effectiveness of foreign aid," *European Economic Review*, 40:2, 289-329; and Craig Burnside and David Dollar, 2000, "Aid, policies and growth," *The American Economic Review*, 90:4, 847-868. These results are reviewed and contested, however, in a recent paper by Michael Clemens, Steven Radelet and Rikhil Bhavnani, 2004, "Counting chickens when they hatch: The short-term effect of aid on growth," Center for Global Development Working Paper 44, <http://www.cgdev.org/Publications/?PubID=130>. Clemens, Radelet and Bhavnani find positive country-level economic impacts from aid based on cross-country econometric studies focusing on the approximately 53% of aid that one would expect to yield short term economic impacts.

institute independent and consistent evaluations of the impacts and cost-effectiveness of donor-funded projects.²⁸ This would mark a significant departure from existing practice, so I first explain the concept, then suggest how it could be implemented, and finally consider how it compares to established evaluation approaches.

For both accountability and learning, the appropriate frame of reference is not the individual project but the donor agency's overall portfolio, and, for learning, the world-wide distribution of similar and near-similar projects. The donor agency's question is how to allocate its resources so as to maximize the impacts of its overall portfolio. The project planner or manager's question is, in light of relevant features of the beneficiary population and the project environment, (and, for managers, in light of how the project is unfolding,) how to configure the project design so as to maximize impacts. In both cases the relevant conception of impacts is one that supports comparisons among projects.

Both accountability and learning, for donor agencies, start from impacts and then work backwards in causality. They start, for example, with strong or weak results, and while accountability uses the discovery of causes to allocate responsibility, learning uses it to construct lessons based on schemes of similarities (so the lessons can be applied to other contexts). In this way rewards and sanctions can be allocated based on contributions to impacts and managers can gain a feel for what is likely to work in a new situation.

²⁸ Impacts are defined as changes in conditions of the beneficiary population due to the project, i.e. compared to the situation one would expect in the project's absence (compared to the counterfactual).

Now this logic may sound quite general. It applies with particular force to large donor agencies because their other sources of accountability are so sparse, the tasks they undertake are so costly and complex, and the contexts in which they work are so often difficult and demanding. Accountability and learning bear a heavier burden than in other contexts. This is why projects should be evaluated not by the extent to which they achieve their individual, idiosyncratic objectives, but in terms of impacts expressed in consistent and comparable units. An evaluation's units of analysis establish a perspective or orientation in terms of which the project and its activities come to be understood. In order to establish a consistent orientation across a donor agency's portfolio, therefore, (or, for a given type of project, across countries and donor agencies,) evaluations should be conducted in consistent units. Accountability requires consistent units precisely because the appropriate frame of reference for accountability is the donor agency's overall portfolio.

3.2 Comparing projects in terms of cost-effectiveness

I would like to suggest that the unit that provides the appropriate frame of reference for donor agency evaluations is cost-effectiveness. Cost-effectiveness aims to achieve the greatest development impacts from the available resources. We can compare evaluation in terms of cost-effectiveness with two other approaches that donor agencies have often used. Bilateral donor agencies have typically evaluated their projects in terms of how far they have achieved their stated objectives,²⁹ and the multilateral development banks (such as the World Bank) have historically evaluated most of their projects in terms of their economic rates of return.

²⁹ This is the 'logical framework' approach.

When projects are evaluated in terms of their objectives, comparisons among projects are likely to be misleading. Some projects have ambitious objectives while others are more modest, so a project that achieves a minority of its objectives can clearly be superior to another that achieves most of its own aims. Also, the criterion “to achieve objectives” bears no clear relation to costs. If this criterion is taken as the basis for accountability, it establishes an incentive to set easy targets and/or to over-budget.

A project’s economic rate of return (ERR) expresses the relation between the sum of the economic value of its benefits and its costs.³⁰ It can also be described as the return on the investment, and the World Bank has typically expected an ERR of 10% or higher from its projects in the economic sectors. The difference between cost-effectiveness as I am defining it and an ERR is that an ERR measures benefits in terms of their economic values (ideally at competitive market prices) while cost-effectiveness measures benefits in terms of the donor’s willingness to pay for them. The economic analysis of projects typically does not include improvements in health or education, and benefits to the poor generally count the same as benefits to households that are already well off.³¹ For an evaluation system based on cost-effectiveness, a donor would need to establish a table of values specifying how much it is willing to pay for each of the various benefits that it expects from its projects, including those that may be expressed in qualitative as well as in

³⁰ Specifically, the ERR is the discount rate at which the discounted sum of benefits minus costs is equal to zero.

³¹ J. Price Gittinger, 1982, *Economic Analysis of Agricultural Projects*, second edition, Baltimore, MD: Johns Hopkins University Press.

quantitative terms. In this way a basis would be established for comparing, for example, primary health care and agricultural extension projects.³²

3.3 The proposed evaluation approach in practice

In practice, the proposed evaluation approach would work like this. At the completion of a project, an evaluator estimates the sum of impacts up to the present point in time and the magnitude of impacts that can be expected in the future that can be attributed to the project. The project's total impacts are compared to its costs based on the donor's table of values, and on this basis the evaluator estimates the project's cost-effectiveness. To estimate impacts the evaluator lists the relevant impacts from a project of the present type and design,³³ and carries out the appropriate quantitative and/or qualitative analysis of the project's activities and their results. The evaluator assigns each form of impact a numeric value and/or a qualitative rating based on his/her judgment of the project's likely effects in the respective areas over the lifetime of the project's influence. The impacts are summed together with appropriate weights from the table of values and compared to costs, and on this basis the evaluator estimates the project's likely cost-effectiveness, for example, on a scale from one to six, with one representing failure and six indicating excellence (see Table 1). The evaluator also notes her degree of confidence in the cost-effectiveness score, and, if her confidence is moderate or low, indicates the range of cost-effectiveness scores in which she is confident that the true value of likely impacts lies. In this case she also specifies the additional

³² This approach to establishing the value of project impacts is described in Paul Clements, 1995, "A Poverty Oriented Cost-Benefit Approach to the Analysis of Development Projects," *World Development*, 23:4, 577-592.

³³ These may be found in the project plan.

information that could plausibly be collected that would allow a more precise estimate. The estimate of the project's cost-effectiveness anchors the evaluator's analysis of the project's design and implementation. All four components—the analyses of impacts, cost-effectiveness, design and implementation—serve as a basic unit to support accountability and learning within the project and the donor agency and across the development community.

Table 1: Scale of Cost-Effectiveness

| Economic Rate of Return | Degree of Cost-Effectiveness | Interpretation |
|-------------------------|------------------------------|----------------|
| 30% and above | 6 | Excellent |
| 20% - 29.9% | 5 | Very good |
| 10% - 19.9% | 4 | Good |
| 5% - 9.9% | 3 | Acceptable |
| 0% - 4.9% | 2 | Disappointing |
| Below 0% | 1 | Failure |

3.4 An evaluation association to address bias and inconsistency

While evaluations in terms of cost-effectiveness may support learning and accountability, there are (at least) three problems with the proposed evaluation approach. First it does not address the bias arising from donor agency control of the evaluation process. Second the estimates of impacts that it requires, including impacts in the future, present methodological challenges. There are no widely accepted methodologies for some of the required impact estimates (e.g. for reproductive health and AIDS education projects). Third, even where accepted methodologies are available (such as economic cost-benefit analysis, for economic projects), the results are often highly sensitive to minor changes in assumptions. When evaluations are contracted out on a project by project basis, different

assumptions are likely to be applied to different evaluations, undermining the validity of comparing and aggregating their results.

There are strong parallels between the conditions for the problem of bias and inconsistency in the evaluation of foreign aid and conditions facing public corporations in the management of their internal finances. Stockholders want corporation managers to employ the corporation's resources in such a way as to maximize profits, but managers face incentives to use the resources for their private purposes. There are elaborate rules governing how managers may appropriately use a corporation's resources, and it is the task of accountants and auditors to ensure that these rules are followed. As with evaluators in foreign aid, however, accountants and auditors are employed by the very managers whom they are expected to hold accountable. In order to protect their independence from management, and to ensure that they have mastered the relevant techniques, accountants and auditors have established professional associations. These associations establish qualifications that their members must achieve and rules that they must follow in order to retain professional membership. It is these rules that are the source of accountants' and auditors' independence from corporate management. Although independence is not maintained perfectly, the consequences of major lapses can be quite severe, as evidenced by the collapse of the international accounting firm Arthur Anderson after the accounts it managed for the Enron Corporation were found to be unreliable.

Amartya Sen lists transparency guarantees as one of five sets of instrumental freedoms that contribute to people's overall freedom to live the way they would like to live. Society depends for its operations on some basic presumption of trust, which depends in turn on guarantees of disclosure and lucidity, especially in relations involving large and complex organizations. Sen points out that where

these guarantees are weak, as they appear to be in foreign aid, society is vulnerable to corruption, financial irresponsibility, and underhand dealings.³⁴ A professional association of development project evaluators could play a role guaranteeing disclosure and lucidity in the management of international development assistance similar to that of associations of accountants in the management of public corporations. Such an association could also address the problems of estimating project impacts and of comparing impacts in common units.

In order to address the problems of bias and inconsistency, such an association would need the same structural features as associations of accountants—qualifications for membership, a set of rules and standards governing how evaluations are to be carried out, and procedures for expelling members who fail to uphold the standards.

In order to ensure that impacts are estimated and then compared in common units, the association would establish a constitutional principle asserting that each end-of-project evaluation conducted by its members would estimate the project's impacts and cost-effectiveness to the best of the evaluator's ability. One task in establishing the association would be to work out impact assessment approaches for different kinds of projects. The technical difficulties in estimating project impacts are objective problems, so it is possible to identify principles and practices for addressing them. An evaluation association would provide a forum for identifying better evaluation approaches and for ensuring consistency in their application. Over time, as its members gained experience, these approaches would be refined.

³⁴ Amartya Sen, 1999, *Development As Freedom*, New York: Knopf Publishers, pp. 38-40.

There are dozens of donor agencies and many thousands of implementing agencies in the development assistance community, and each agency has its own management culture and approaches. The universe for the evaluation association's operations would be the development assistance community overall, and it would support learning about better practices on the basis of the type of project throughout this community. Each evaluation completed by a member of the association would be indexed and saved in an online repository, which would be accessible to the entire development community. Since each evaluation estimates the project's cost-effectiveness, it would be a simple operation for someone planning, say, an urban water project, to review the approaches of the five to ten most cost-effective water projects in similar environments.

4. Monitoring and Evaluation (M&E) for Cost-Effectiveness Compared to Other M&E Approaches

4.1 Monitoring and evaluation for empowerment

To explain M&E for cost-effectiveness it is useful to compare it with other evaluation approaches. The strongest challenge to standard approaches to aid evaluation in the last two decades has involved the elaboration and application of participatory approaches.³⁵ These have aimed to involve beneficiary populations in project management, to assist them in taking responsibility for improving their own conditions and to incorporate them in more democratic processes of development decision making. Authors such as Korten and Chambers,³⁶ whom

³⁵ See e.g. B. E. Cracknell, 2000, *Evaluating Development Aid: Issues, Problems and Solutions*, Thousand Oaks, CA: Sage Publications.

³⁶ David C. Korten, 1980, "Community Organization and Rural Development: A Learning Process Approach," *Public Administration Review*, 40:5, 480-511; Robert Chambers, 1994, "The

Bond and Hulme describe as “purists,”³⁷ have sought to reorient the development enterprise to support the goal of empowerment. They have promoted an approach I call “M&E for empowerment” because it emphasizes learning at the local level, seeking to empower project beneficiaries by involving them in the evaluation process. While M&E for cost-effectiveness appreciates that empowerment is an important development goal, it identifies the locus for the primary learning that evaluation should support among those who are responsible for resource allocation decisions. Donor agency officials are the primary audience for aid evaluation because they exercise primary control over these resources. It turns out, however, that the form of evaluation that can best inform these officials will also best inform officials of developing country governments, project managers, and the overall development community, as well as, with some additional synthesis, the legislatures that appropriate aid budgets.

Evaluation and empowerment goals overlap in their management implications, and empowerment was certainly neglected by the development community prior to the mid-1970s. In many instances participatory strategies are more cost-effective than projects based on so-called blueprint approaches, so M&E for cost-effectiveness would promote participation in these cases. M&E for cost-effectiveness does not assume, however, that participatory approaches are right for all projects. The empowerment of project beneficiaries is interesting from an analytic viewpoint,

Origins and Practice of Participatory Rural Appraisal,” *World Development*, 22:7, 953-969; Robert Chambers, 1994, “Participatory Rural Appraisal (PRA): Analysis of Experience,” *World Development* 22:9, 1253-1268; Robert Chambers, 1994, “Participatory Rural Appraisal (PRA): Challenges, Potentials and Paradigm,” *World Development*, 22:10, 1437-1454.

³⁷ R. Bond and D Hulme, 1999, “Process Approach to Development: Theory and Sri Lankan Practice,” *World Development*, 27:8, 1339-1358, p. 1340.

because it can be seen both as a means to improving project designs and as an end in itself. For this reason M&E for cost-effectiveness views empowerment in a dual light. As a means, M&E for cost-effectiveness considers empowerment like any other possible means to be considered in program design. As an end, M&E for cost-effectiveness considers successful empowerment to be a benefit which must be valued and counted along with other benefits in the assessment of a project's cost-effectiveness. Under M&E for cost-effectiveness both more and less participatory projects are considered within the same evaluation framework.

4.2 Monitoring and evaluation for truth

It is possible that a great practical barrier to useful evaluation arises from some of those most knowledgeable of and committed to evaluation as a science. It has been common practice to begin discussions of aid evaluation methodology with the experimental method of the natural sciences,³⁸ and to present the various evaluation methods as, in effect, more or less imperfect approximations to randomized and controlled double-blind experiments. This approach often uses household surveys that measure conditions that a project seeks to influence, so that through appropriate comparisons changes attributable to the intervention can be identified in a statistically rigorous manner. I call it "M&E for truth" because it emphasizes making statistically defensible measurements of project impacts. This approach is right to insist that projects should be assessed primarily on the basis of their impacts, and that impacts should be understood as changes in the conditions

³⁸ E.g. Dennis J. Casley and Denis A. Lury, 1982, *Monitoring and Evaluation of Agriculture and Rural Development Projects*, Baltimore, MD: The Johns Hopkins University Press; Judy L. Baker, 2000, *Evaluating the Impact of Development Projects on Poverty: A Handbook for Practitioners*, Washington, DC: The World Bank.

of the population compared to what would be expected in the project's absence (in evaluation jargon, as compared to the counterfactual). It is arguable, however, that in its orientation to statistical rigor it has established a "gold standard" that many evaluators are all too quick to disavow. Only a very small proportion of project evaluations present statistically rigorous impact estimates, and evaluations that do not often use the demanding requirements of statistical rigor as an excuse not to address the question of impacts at all. Also, evaluations that adhere rigorously to the maxims of statistical rigor seldom estimate the future impacts that can be attributed to a project.

Monitoring and evaluation for cost-effectiveness is methodologically eclectic in its effort to reach reliable judgments of cost-effectiveness. It is grounded not in the first instance in the scientific method, but in the causal models of change inherent in project designs. Each project design presents a hypothesis as to the changes in beneficiary conditions that can be expected from the actions the project undertakes. It is the evaluator's task to analyze how this hypothesis has unfolded, and on this basis to estimate the quantity of benefits that beneficiaries are likely to realize. A given project is taken as an instance of a project of its type, so impact estimates for other similar projects serve as a first approximation for the benefits that may be anticipated from the present project. Evaluators locate the present project along the continuum established by other similar projects based on how its design hypothesis unfolded as compared to theirs. Clearly, baseline surveys often provide critical information for estimating impacts, and statistical methodology of course provides central criteria for their analysis. As suggested above, M&E for cost-effectiveness employs participatory methodologies in many instances to elicit beneficiaries' judgments of the significance of project outputs. The evaluator's final estimate of a

project's impacts and cost-effectiveness, however, is based on triangulation taking into account all the forms of information we have so far considered.

5. A More Analytic Development Assistance Community

Although I have described the proposed approach as monitoring and evaluation for cost-effectiveness, the discussion up to this point has focused on evaluation only. For the proposed evaluation approach to address the accountability problem in foreign aid, however, it is essential that planners and managers should know in advance that upon its completion there will be an independent evaluation of their project's impacts and cost-effectiveness. The development assistance community as we know it has evolved under conditions of inconsistent and often limited and biased evaluation, but one could anticipate, if the proposed evaluation approach were implemented, that its effects would gradually suffuse through all stages of project planning and implementation. Project planners would soon learn to include an estimate of cost-effectiveness in their project designs, and to establish monitoring systems that would track the relevant impacts (or their main contributing factors) through the life of the project. It would soon be taken for granted that when targets or systems for estimating impacts are altered during project implementation, the reasons for these changes should be clearly documented. The development assistance community would soon learn what outcomes need to be tracked for different kinds of projects to inform subsequent impact estimates.

While many individuals and groups in the contemporary development community are engaged in promoting development agendas of their own conception, the proposed reforms would enhance the experience of development work as a cooperative venture with shared goals. Development professionals would become

more confident that others would endorse their sound justifications for their management strategies, and management strategies would be more rigorously grounded in expected impacts. Members of the development community generally would become more conscious of the pathways by which their actions contribute to improvements in beneficiary conditions, and their shared concern for efficiency would be enhanced. The development community would be quicker to identify and to adopt more successful strategies. Although I believe that outright corruption on any significant scale is uncommon in the development community,³⁹ the higher analytic standard that the proposed reforms would bring about would reduce corruption even further.

The general public has tended to be fairly skeptical of foreign aid, and the management standards described in this article provide good reasons for skepticism. The proposed reforms would make it straightforward to aggregate project impacts, for example, by country or by agency. The tax-paying public would receive better information on the consequences of foreign aid, and they would have better grounds for confidence in its integrity. In due course this could be expected to increase the generosity of citizens in the wealthier countries towards people in need.

³⁹ I found large scale corruption in only one of the 12 projects in the sample for my dissertation.

Network Evaluation as a Complex Learning Process

Susanne Weber⁴⁰

The following contribution will explicate, based on an understanding of networking as a reflexive process and on an approach working from a theory of regulation, one set of criteria for the development of evaluation designs in a networking context. Needs for evaluation and monitoring that is action- and future-oriented lead to other needs already established by social-ecological planning theory. From these can be generated questions for decisions in monitoring and evaluation within complex actor settings as well as criteria for concepts of evaluation and monitoring in a networking context. On this basis, four dimensions of network evaluation and monitoring are suggested and they are embedded in the multi-dimensional design approach of the “learning network” which puts collective competence development and future- and effect-orientation at the center of the developmental process.

1. Networks: Between myth, management and “muddling through”

Networks are by now being discussed in all disciplines of the social sciences as the new paradigmatic form of organization and pattern for action. There are divergent assumptions about their status and range of applicability, their application contexts can be political, economic or social, and applications serve numerous possible

⁴⁰ Corresponding author: Susanne Weber, PH.D. University of Applied Sciences, Fulda, Department of Social Studies, Marquardstr. 35, D-36039 Fulda, Germany, telephone: 0049-661-9640-224, e-mail: webers@mail.uni-marburg.de or susanne.weber@sw.fh-fulda.de. Paper prepared for European Evaluation Society Sixth Conference, Berlin, 2004

networking goals and purposes. For these reasons the term “network” is defined as a “compressed term” (Kappelhoff 2000:29): networking represents a perspective of hope, a factor conducive of democratization and successful cooperation, professional optimization, rationalization, market presence, and as a term employed almost as universally as the term “system” (Grunow 2000:314), it is often very nearly mythologized (Hellmer/Friese/Kollros/Krumbein 1999). It is used to represent a variety of possible meanings and forms of cooperation with different degrees of intensity: following Simmel (1908), society for instance is now once more increasingly explained in terms of network theory (Castells 2000; Messmer 1995; Wolf 1999), with “network” as one of the basic social categories.

Networking is also the point of departure for more or less close forms of cooperation in a regional context, often initiated by support programs and generating research interest in practice and action, e.g. the “Learning Region” program. Here we are dealing not only with a clear accentuation of the term “network”, but with a “school of thought, a line of orientation, a ‘warmth metaphor’ including an accentuated demand for initiative: regions shall be guided out of their passive role, taking on an active part in dealing with their concerns” (Gnahs 2003:100). As part of regional networking processes, intermediary agencies for regional learning networks are created which are supposed to tie different social fields together, to give creative support (Jutzi/Wöllert 2003:130) and to serve as bridges for the initialization of regional processes by defining needs, giving orientation, maintaining and integrating patterns (ibid.:135).

Sydow suggests a tighter definition, and thus a higher degree of intensity for networking, characterizing it from a micro-economic perspective on company networks as “a form of organization of economic activity by enterprises which are independent by law but more or less interdependent economically”. The relations

that are introduced here are reciprocally complex and rather cooperative than competitive. They are relatively stable, they are created endogenously or induced exogenously and represent more or less “polycentric systems” (Sydow 2001:80). They can be categorized e.g. by their type of control and the stability of their relations (stable-dynamic) (ibid.:81). For applications in competence development Duschek and Rometsch (2004) suggested grouping the various network types into three main types: explorative versus exploitative, hierarchic versus heterarchic, and stable versus dynamic networks (ibid.:2).

Risk and conflict are inherent to the structure of such institutional and organizational cooperations (Messner 1995). Due to their structural complexity, they are not always at an advantage over other forms of organization: Steger (2003) identifies contradictions like self-interest versus collective interest: in building common structures of action, a chance for creating common space for development curtails the flexibility of individual network actors; the commitment that becomes relevant in a networking context reduces the autonomy of the individual network partners, etc. (ibid.: 13f). Sydow (1999) presented the model of structural tension in network cooperation, which will be further discussed below since it can be made productive for the analysis and design of monitoring and evaluation in network cooperations.

The theoretical framing of the network and network arrangements proves to be of decisive importance for the design of network cooperation as well as for the evaluation-theoretical and conceptual position of monitoring and evaluation in a network. In this paper, network cooperation is discussed on a social-scientific basis, as a social process in which the surfacing of specific conflict potentials, risks, and tensions, is to be expected. Theoretical perspectives that have complexity (Kappelhoff 2000) and structuring (Sydow 1999; Windeler 2001) as

their starting point are capable of representing and analyzing this topic in a way that is adequate for design practice and network management at the same time.

2. The approach of network regulation as a theoretical foundation for monitoring and evaluation in networks

The network regulation approach offers criteria for the conceptual level of monitoring and evaluation in networking contexts. The five characteristics of network regulation show us consequences for the design of monitoring and evaluation.

Constitution

One aspect of the five characteristics that belong to an approach following a theory of structuring is a procedural understanding of constitution, which transcends the static look at organizational networks. The network constitutes itself in time and space via social practices, as a collective social setting. It regulates itself systemically and contextually (Windeler 2001:203f). From the perspective of a theory of structuring, monitoring and evaluation are not outside the networking activity, they are part of the system and are systemically generated by it. In the context of a regulatory system monitoring and evaluation are also regulated and constitute themselves during that process.

Multi-dimensional regulation

The existence of different levels of actors is characteristic for networks: that of the individual, the group, the organization, the network itself and society as a whole. Multi-dimensional regulation means that divergent interests on the different levels are regarded as structurally unavoidable. The different levels of actors become relevant for the employment of complex monitoring and evaluation in networking

processes. One has to deal conceptually with the question which levels should be included for the generation of knowledge and how, and what consequences are intended. One has to ask what different goals and goal achievements in a multi-level context are to be analyzed and how multiple goal structures figure in monitoring and evaluation designs.

Contextualization

The third assumption of a regulatory approach is that the constitution of organizational networks is a coordination of activities in time and space. Networks are embedded in specific contexts and environments that play an important role for conditions and cultures of action. Every network will develop its own context-specific culture and specific social memory (Windeler 2001:325).

Network monitoring and evaluation will be designed, and will have to be designed, according to the respective network culture. Thus we can distinguish sector-specific evaluation cultures: In the profit field we find a strong orientation toward planning while networks close to the administration, which may e.g. be confronted with a need to legitimize their activities because they receive public funds, rely on summative and ex-post evaluation. When evaluation concepts are dealt with according to sectors, this then includes practice oriented toward planning and resources, toward process and correction, or toward summative legitimization.

Co-evolution

From a theoretical perspective of structuring—and this is the fourth aspect—the development of organizational networks can be seen as a process of co-evolution with the relevant environment. Co-evolution means that context relevancy cannot be ignored, that the embedding in institutional contexts and relevant environments

has to be considered. Not only the inner core of the network which is to change, but the participating organizations as well are exposed to change, so network monitoring and evaluation are capable of fulfilling a learning and development function for the inner core as well as for its environment. It remains an open question in each case to what degree the collective actors are able to reproduce their system reflexively and to establish reflexive monitoring. Subcultures, subgroups, subunits will describe and interpret themselves and their respective situations differently. System monitoring can establish practices which throw a light on the experiences and expectations of the network partners (Windeler 2001:326) and which co-evolutionarily reconnect the environment to the system's inner workings.

Networking in terrains structured by dominance

The fifth aspect of an approach according to a theory of structuring is recognition that organizations as collective actors interact competently and powerfully on a terrain structured by dominance (ibid.:30ff). Network membership is very intentional, discursive, strategically important and available (ibid.:251). Evaluation shows a very sensitive relation between contribution and use, and in antagonistic settings it can be contested between stakeholders. That is why procedures and programs need to be designed that analyze the contributions and potentials of the individual actors, the practices and activities as well as the networking context as a whole. General criteria for evaluation and responsibilities for their design, for monitoring, for compliance with these criteria, and for sanctioning, have to be developed reflexively.

Network monitoring and evaluation face the challenge to analyze not only factual dimensions like the management of business activities, but also potential power-

driven roadblocks in dominance-structured fields of action, e.g. veto and blocking positions, minimal consensus in goal definition, the curtailing of autonomy of network partners, refusal to learn, and the shifting of risks onto third parties (Sydow 1999:298). The evaluative function (ibid.) of network management is supposed to put the whole range of social, factual, and procedural aspects of network management to the test.

Sydow bases the relationship between network management and network development on a theory of structuring (2001:82): network development is seen as observed change over time within a social system that is reproduced by relevant practices. Change takes place in a planned way through intervention and also in an unplanned way, through evolution. This perspective relates to the process by which network actors refer to network structures in their actions and attempts at guidance, reconstructing those structures by their actions (ibid.:83). Incorporated in this are structures, ways of development, and the possibilities of trans-organizational development—but also the possibility of failure, of unintended results, of alternative actions, of coincidence. Network development (and the effects and feedback effects it has on the organizations involved) can be described as the result of reflexive as well as non-reflexive structuring (Windeler 2001).

To make networks more successful—and this is the procedural and future-oriented function of monitoring and evaluation that will be the center of our attention here—it makes sense to analyze and to design network management as reflexive network development. Monitoring and evaluation then gain central importance for network development: They facilitate the understanding of network development as a field of learning and of the collective development of competence (Weber 2002, 2003, 2004), and they suggest the importance of analyzing empirical networking projects (Weber 2001a).

What then should concepts and designs for network evaluation and monitoring look like? This question leads to others, common in evaluation settings, e.g.: What information shall be generated, and how? What knowledge is needed and functional? What function should reflexivity have, what should it achieve? Who should generate knowledge and what should it be used for? If we take the program seriously that was elaborated at the 2003 DeGeVal convention—evaluation should lead to organizational development (Hanft 2003) and the focus should be shifted from summative and ex-post analysis towards process monitoring and future development (Weber 2001b)—then it makes sense to follow the incremental theory of planning. The social-ecological theory of planning and the 1970s’ criticism of classical theories of planning give us criteria that can be used for the reflection and conceptualization of evaluation designs.

3. Selection decisions for the generation of knowledge in networks

Uncertainty about the individual actors’ judgement, the comprehension of the original situation, the actors’ collective action, future developments and strategies under a perspective of transformation (Schäffter 2001) can be made productive, if tied to monitoring and evaluation in a view supported by a defensible theory of science.

Following an objectivist or constructivist understanding of reality we can distinguish an “objectivist” from a “constructivist” evaluation paradigm. These different understandings will now be considered in their polarity, and afterwards their functionality for monitoring and evaluation in a networking context will be discussed. Their polarity brings monitoring and evaluation into focus as not just instruments, but as networking practices.

While in classical concepts (Rossi/Freeman 1989:18) evaluative approaches were regarded as analytical instruments without the ambition of serving as theory-guided science (Kuhlmann 1998:92), we here consider concepts and approaches to monitoring and evaluation as active practice which is part of and generates specific network cultures. We assume that settings for communicative evaluation are not “just instruments”, without preconditions and “objective”, but that in reality they have a generative quality, organizing observation and knowledge production according to underlying explicit or implicit criteria and models of evaluation. Working on organizational transformation processes, Roehl and Willke have pointed out the—often substantial—“constructedness” of evaluation settings, which is brought about by the choice of instruments and criteria. Evaluation designs are always subject to leading ideas of change, which include ideas about the validity of changes and which, in a context of complex structures of decision-making, predetermine the evaluative direction (Roehl/Willke 2001:29).

Drawing on cybernetic, social-ecological or systemic criticism of planning in the 1970s and 1980s (Lau 1975, Atteslander 1976) decisions can be identified that become relevant to the selection of evaluation designs. E.g. in the 1970s’ criticism, dimensions of subjectivity, communication, and system orientation are emphasized in the face of a rationalist, technocratic paradigm. This criticism leads to an alternative planning paradigm that includes choices that are relevant for the design of planning and monitoring, such as the following:

- Between “technocratic feasibility” and “systemic irritation”
- Between legitimization of the past and planning of the future
- Between the reproduction of the old and the generative production of the new

- Between “expert objectivity” and subject participation
- Between the completeness of what is known and the processing of what is not known/uncertain
- Between result measurement and the development of competence

These selective decisions can be found, in different manifestations, in today’s evaluation practice in different social contexts, and their range, their deficits, and their chances for “reality construction” can be analyzed. The following presentation of decisions and questions relevant to evaluation in network settings does not pretend to cover all aspects comprehensively; instead it treats them by means of examples.

3.1 Evaluation knowledge between “systemic irritation” and “technocratic feasibility”

Within the evaluation community there is a tension between two contradictory approaches, either of which follows from basic questions of a theory of planning. A “technocratic” approach builds on the assumption that existing knowledge can be used to give an intentional design to social conditions (Herrmann 2001:1365), that social processes can be rationally planned and influenced. On the other hand there is the contrary view, skeptical of a teleological regulative approach to social processes that presupposes predictable results. This view assumes that even the most advanced and differentiated instruments of planning eventually cannot “handle” social reality.

In the 1970s, models that take an optimistic view of regulation are increasingly opposed by regulation-skeptical models calling for more open and dynamic approaches to planning and evaluation. Early on, Lau (1975) pleads for

management of complexity through a participative concept of planning that retains a sense of flexibility. Atteslander presents a typology of different planning models and defines a dogmatic, a technocratic and a cybernetic or systemic type (Atteslander 1976:20).

The systemic-constructivist assumption of the self-organization of institutional systems leads to a concept of planning and thus of measuring effectiveness and evaluation which is based rather on “irritation” than on technocratic “feasibility”. Reflexivity is encouraged and facilitated in order to partially produce uncertainty (Herrmann 2001:1365).

3.2 Evaluation knowledge between legitimization of the past and planning of the future

Another dimension pertinent for today’s evaluation debate is the directedness towards past, present or future. Evaluation or monitoring designs aim, to varying degrees, to create legitimacy or change and complex transformation. The directedness of evaluation designs towards past, present or future is today influenced by sectors and organizational cultures.

An “evaluation culture” in the sense of a summative evaluation emphasizes the thorough analysis of the past, the evaluation of previous projects. Here the aim is often legitimization, and evaluation is rather geared towards a bureaucratic model of control, transparency, and the evaluation of goal attainment. The focus is set on the summative evaluation of individual measures and programs without strong references to organizational visions and goals, and the activities are relatively little strategically synchronized or planning-oriented.

A “monitoring culture” on the other hand emphasizes a process-accompanying, formative evaluation and self-evaluation. Goals connected to monitoring and evaluation are endogenous development, motivation instead of control, process-orientation, and improvement on the level of professional action. Possible risks lie in conducting many parallel activities on all levels (supervision, etc.) which do not receive feedback from each other, which are not directed toward the organizational or networking goal, and which see themselves as strategically oriented. A tendency towards monitoring with self-evaluation classically corresponds with the evaluation concepts preferred by the non-profit sector.

Evaluation designs which are more strongly embedded in a “planning culture” emphasize diagnosis, feasibility studies, and conditions for success; they do not rely very much on summative evaluation. Their focus is on future orientation, financial aspects of a cost-benefit relation, numbers and control. The most effective interventions harmonize with visions and strategies of the system of reference, in this case the network. The aim is not the realization of individual activities but the strategic feedback relationship of all measures that is supposed to create an equal directedness of all activities.

3.3. Evaluation knowledge between reproduction of the old and generative production of the new

There is also a tension in monitoring and evaluation between the reproduction of the old and the generative production of the new. This tension is already implicit in the demands made during the planning debate of the 1970s: instead of mechanistic models for planning, the generative production of the new was to be facilitated. Instead of prognoses of the future based on the status quo, “anticipation” was to be employed systematically. The inclusion of prophecies and projections of all kinds

in a context of cybernetic models of planning was seen as more adequate to the challenges and demands of planning than dogmatic or technocratic models of planning (Atteslander 1976:53).

3.4. “Expert objectivity” or subject participation

The fourth decision in evaluation represents the distance between evaluation by experts and by participants. Evaluation by experts is often oriented at utilitarian-rationalist models of action and leaves responsibility in the hands of the expert. The participants tend to become objects of the evaluation, not systemic partners in collective efficiency measurement and evaluation.

In a heterarchic decision-making structure, democratized expertise is a given and the production of knowledge that becomes relevant for action has to work with network knowledge—if it does not, there are distinct risks of interest-guided dominance and colonization on the one hand, lack of acceptance and inner emigration by network partners on the other. Knowledge production in networks thus has to rely on the cooperative structures of “participatory research” (Atteslander 1976:53). The efficiency of the solution of material problems depends on the participation of those concerned, on openness to criticism, on horizontal structures of interaction and on democratic procedures for implementation.

3.5 Completeness of what is known or processing of what is not known/uncertain

This decision is tied to different accentuations—is knowing or not-knowing the point of reference? Open and dynamic models of planning and of monitoring assume a transformable worldview and a comprehensive definition of goals. They do not presuppose knowledge but rather incomplete knowledge or no knowledge

about the current situation and its structures. These approaches are synthesizing—they methodically attempt to integrate ideological, technological and social aspects of networking contexts.

The rationality and the kind of prognoses connected to cybernetic efficiency measurement and evaluation can be described as an operational rationality working with a combination of deductive and normative prognoses. In an incremental view, planning can never be final, it is always preliminary and influenced by a large amount of feedback (Atteslander 1976:55). It systematically needs monitoring and evaluation.

Monitoring and evaluation are more than social technology in this case, they are reflexive practice and the creation of communicative contexts where the constitution of social meaning takes center stage. The focus is not on needs presumed to be objective, but on the needs and perspectives of the network actors.

3.6 Result measurement or the development of competence

Contrary to a basic view of processes of planning, monitoring and evaluation as “technology”, an understanding of planning and monitoring as something to be negotiated is directed towards the development of competence. Contrary to placing planning, monitoring and evaluation before or after actual practice, an integrated view is suggested, which shifts away from a purely concept-oriented evaluation of efficiency towards one that also considers (micro- and meso-) political structures. Monitoring and evaluation are no longer primarily goal-oriented, instead the area of work becomes evaluation-oriented. Measurement of efficiency and evaluation can tune in to a daily networking routine that changes slowly. This understanding goes hand in hand with an increase in competence and with self-rationalization of the network partners. By taking into view the social aspects of the production of

knowledge that is relevant for implementation, an open model aims at the development of competence, functionality in poly-centric and heterarchic structures, and the internal democratization of expertise (Atteslander 1976).

Communicative planning concepts are process-oriented, not schematic; they follow the principle of negotiation. They do not pretend to be neutral in terms of values but facilitate working through the topic of value, the equivocal connection between ends and means in social contexts. Communicative approaches are the only planning approaches that attempt to bridge the gap between conceptual planning and practical action by conceptually integrating the problem of the implementation of planning results. Communicative planning practice further documents that such models represent adequate concepts for action within the complex and contradictory conditions and processes in areas of planning (Herrmann 1998; 2001:1378).

These short sketches of considerations based on a theory of planning can be used for the design of instruments and concepts of evaluation and monitoring. They address central questions about the basic assumptions, the direction and starting points of analysis, about the status of evaluation in networking contexts, implicitly also about instruments and procedures, and they furnish a pattern for meta-evaluation, in so far as evaluation concepts themselves become objects of evaluation with the help of certain criteria.

3.7 Consequences for the design of monitoring and evaluation in networking contexts

It has become evident that classical evaluative approaches reach their limits in networking contexts. E.g. in complex program evaluation it could be shown that the fact has been neglected that programs follow “multiple, conflicting and

evolving purposes” (Kuhlmann 1998:97), that the context of their conception is often not sufficiently understood, that evaluation is used as a “killer”, that the views of those who are responsible for the program are taken into consideration but not the interests of those concerned (ibid.:98). In the context of a multi-layer concept it is neither possible nor sensible to measure “objective results” exactly, in the sense of eternal truths (ibid.:85). Under a perspective of reflexivity in a networking context, communicative validation, process monitoring and evaluation become integral parts of network regulation as a design approach.

A social-ecological planning paradigm becomes manifest, demanding a mainly communicatively oriented validation, an incremental communicative practice of planning and action that is more adequate to the necessities of the field than classical evaluation designs (Zipp 1976:77). These demands can be tied to the social-ecological approach to evaluation developed by Guba and Lincoln (1989), continued in participative approaches to evaluation (Ulrich/Wenzel 2004) and implemented empirically (Uhl/Ulrich/Wenzel 2004).

Networks are exposed to structural uncertainty about the future, and in their intended and unintended reflexive practice, in the systematic form of evaluative, process-analytical and planning practice with a perspective of collective development of competence, they can be reconstructed as “learning” networks (Weber 2002), i.e., as a field of pedagogical rationality (among others) (Helsper/Hörster/Kade 2003). Intended and unintended qualities of learning will find their space here. Informal, quasi-evolutionary learning processes as well as orchestrated reflexive interventions can generate learning and reflexivity in a “learning network”. Learning (on the different levels of individual actors, groups of actors, the network structure and its relevant environment, up to the social body as a whole) is contingent and uncertain. As learning from experience it is

intertwined with everyday working activity. If the knowledge-generating practice of making experiences “on the job” gets established, systemized, and structurally put into a feedback relation with the system’s practice, then orderly procedures for institutional and network learning are created. Learning on the subject level and on different system levels also becomes systemized, and monitoring and evaluation are put into a context of a development-oriented strategy of collective learning within the network. Depending on to which dimensions the reflexive generation of knowledge within the network can be designed around, they will be the focus of the following section.

4. Dimensions of evaluative and planning-oriented learning within a network

On the basis of empirical networking projects (Weber 2001a) and literature on networking theory we can determine four dimensions for a strategy of collective learning. These design dimensions of system monitoring and evaluation are the social dimension, the dimension of network functions, that of structural tensions created by networking processes and that of learning and learning arrangements (Weber 2002, Weber 2003). In our analysis of network functions and structural tensions we follow the works of Sydow (1999, 2001) and Windeler (2001). While these two aspects have already been objects of network regulation as well as reflexive approaches, the dimension of the social process and that of learning have not yet been considered systematically under a design point of view.

4.1 Social regulating and social monitoring

The regulatory approach presupposes structuring by social actors, so the social dimension is thus structurally included. Cooperative inter-organizational relations are seen as based on social processes; personal and social closeness is regarded as a necessary condition for successful networking processes (Winkler 2002:37).

Network knowledge is always social, it is created by and embedded in social practice, with its individual and collective elements. As a whole, network relationships are based on exchange, which is, in turn, based on stable expectations and a norm of reciprocity. Trust is also seen as a *sine qua non* for successful projects (Windeler 2001). This shows that the social dimension is indeed recognized as relevant, but so far it has not been addressed in its quality as a group context. Supported by group-dynamics and team-development approaches, we can here refer to categories of the group process that Tuckman uses to describe different group and team qualities (1965). Tuckman's model assumes that the first phase of groups' encounters is friendly and noncommittal, while the second phase of the process is characterized by struggle for social status and power within the social structure. In a third phase the group then has to come together to a functional whole, and positions in social space have been negotiated. As a fourth phase we see the "performing group". Tuckman extends the group, capable of working and performing, into the future.

Theoretical positions based on structuring and complexity describe networks as co-evolutionary entities (Kappelhoff 2000b:382) that do not show a linear development. Still, Tuckman's four-phase approach is useful for its qualitative criteria for the analysis and design of group-dynamical aspects. His definitions of Forming, Storming, Norming and Performing in a social context can be used for monitoring and for evaluation since they provide criteria for the analysis and design of the social context which can be found empirically with the help of indicators.

4.2 Functional network guidance—monitoring of network functions

Another dimension of monitoring and evaluation is the functional dimension of network guidance introduced by Sydow (Sydow 1999). All elements of network regulation—selection, allocation, evaluation, system integration, configuration of positions, constitution of borders—can be objects of evaluation: the selection of the actors belonging to the system, the allocation of resources, the evaluation of the process and the specifics of system integration, and of the configuration of positions and the constitution of borders (Windeler 2001:249).

In a design approach, “selection” includes the question of “who?”—who shall be included? This question becomes important at an early networking stage. After that the focus shifts to the “allocation” of tasks and resources, the distribution of responsibility among the partners. “Regulation” of cooperation within the network provides the development and implementation of rules between the organizations. “Evaluation” of network organizations can concern the network as a whole or just selected rules of cooperation (Sydow 1999:295f).

Windeler adds two others to these four functional aspects: “system integration” and “border management” (Windeler 2001). Measures of system integration influence the selection of actors; the practice of configuration of positions and of the constitution of borders pose particular challenges to potential newcomers etc. (ibid.:251).

These objects of network regulation are interconnected in a recursive relation. The six aspects of network guidance are open to analysis and elaboration under the focus of a functional dimension. While Sydow describes them as procedural, they do not just develop their relevance along the stages of the process but also across them: selection, allocation, regulation, evaluation, border management and system integration are necessary and have to be repeated perpetually and circularly. They

offer a catalogue of questions, criteria and indicators for network monitoring and evaluation along the emergence of design necessities.

In Sydow's approach to network functions (1999:298) monitoring and evaluation have systematic value. The characteristics of reflexive network regulation provide a concrete basis to the function and design of monitoring and evaluation. All in all it becomes evident that network monitoring and evaluation have to be integral parts of a complexity-oriented reconstruction of networking processes. Sydow assumes that monitoring and evaluation become important factors in the design of paths of development within reflexive network development. They furnish the informational basis for a (more) reflexive network development by network management. While "evaluation" aims at the contributions of individual network organizations, at the quality of the network relations that have been developed or at the "network effect", and while as a function of management it is concerned with the practice of evaluating, "reflexive monitoring" is designed as a tool for supervision of one's own actions, of the conditions and effects of actions and of the actions of others (Sydow 2001:90). From a design perspective, monitoring and evaluation facilitate the systematic regulation of networking risks and the increase of networking success.

4.3 Structural tension—monitoring of tension

A third focus of complexity-oriented network monitoring has to be the dimension of structural tension. Sydow has introduced eight lines of tension that have to be regulated in networking processes—or if lacking regulation can cause a networking process to fail (Sydow 1999). They provide analytical potential and differentiating criteria for the evaluation and design of network cooperations. Messner, coming from political science, has also identified structural dilemmas of

networking that have to be worked on within networking processes (Messner 1995, 1994). The following section is based on Sydow's presentation (1999) of the lines of tension between "autonomy and dependency", "trust and control", "cooperation and competition", "flexibility and specificity", "variety and unity", "stability and fragility" (e.g. change), "formality and informality", "economic rationality and preservation of power" (Sydow 1999:300).

Variety—unity: How can a balance be reached between the variety of participating actors and their integration to some kind of unity?

Flexibility—specificity: How flexible is the network in terms of its goals and self-image, how specific is it?

Autonomy—dependency: How much autonomy is possible and what does it consist of, how much dependency exists and what does it consist of?

Trust—control: How much trust and what kind of trust is there; what is regulated by control mechanisms, and how?

Cooperation—competition: What role do cooperation and competition play? What relationship is created between them?

Stability—fragility: What role do stability and fragility play? How are they created? What regulating mechanisms exist?

Formality—informality: How is the relationship between formality and informality regulated, what relationship do they have?

Economy—power: What relationship is there between arrangements of functionality and power? How are power patterns generated?

Windeler (2001) also refers to these lines of tension in his approach based on a theory of structuring. Within a monitoring approach they can be regarded as analytical dimensions and as design parameters. They are useful for the incorporation of reflexivity in discursive and qualitative processes of analysis, thus for clarification and localization within the discursive context and the network's path of development.

4.4 Knowledge, communication and system reflexivity: networking as a learning process

Since networks represent dynamic rather than static arrangements of relations and cooperations, networking has to be read as a learning process. Monitoring and evaluation have the function to generate knowledge from practical experience and to reflect on it, in order to deduce knowledge from it that may guide future actions (Uhl/Ulrich/Wenzel 2004:11). So their primary objective is to provide chances for learning and optimization on the system level. They are tied to system reflexivity and communication, re-entering into the circle of active planning within the network. The explicit directedness of monitoring towards the design of learning contexts makes it possible to identify future-oriented developmental potentials of networking projects. Discursive reflection produces awareness of change in the first place—data gathering procedures not only reconstruct their subject in different ways, the subject of reflection itself is changed by it (Hendrich 2003:157). In network contexts as informal learning contexts, aspects of a learning biography and the estimation of one's own competence can also be used for a kind of monitoring that is oriented to competence development.

The social dimension, the functions of network guidance, the structural lines of tension, and the dimension of learning within the networking process, have been

suggested as dimensions for the monitoring of efficiency and for the evaluation of complex transformations (Weber 2003). What instruments and learning arrangements can support complexity-oriented monitoring and evaluation which reply to demands on the social, functional, structural and learning dimensions in a pragmatic and manageable fashion?

5. A perspective: instruments for evaluative and planning-oriented network development

As the criticism of under-complex evaluation designs has shown, the focus may not be narrowed to a few efficiency indicators since this includes the risk of distortions. Especially quantified data is often endowed with a status of objectivity that makes it difficult to question the results. Under-complex designs for monitoring and evaluation have counterproductive effects when the truth production of the system generates faulty attributions and labeling or unintended effects, e.g. in the sense of social dynamics. This means that monitoring and evaluation in a network have to be geared towards communication and complex reconstruction.

In complex social systems reflexive network monitoring will not exclusively be left to process counselors, brokers, coordinators and moderators. It will be part of everyday action and has to be functional in terms of the necessities that come with this. Below the level of external evaluation by experts it is recommended that there be developed a discursive, procedural self-evaluation. On this level networking needs “cooperative core competence” to balance existing tensions. These tensions cannot be dissolved; they are part of the structural characteristics of networking and have to be dealt with productively. In this way they become accessible to process evaluation and optimization. Sydow thinks that a continuous employment

and practice of reflexive monitoring can render more formal evaluative methods unnecessary (Sydow 2001:97).

Monitoring and networking in networks are instruments for the construction of reality, wrapped up in a heterarchic and polyvalent structure of interests and in complex transformation processes. For an integrated design of monitoring and planning it seems to be practical to generate open evaluation designs (Lynen von Berg/Hirseland 2004:15). These designs should take the form of participative evaluation (Oels 2003, in publication; Weber 2003; Weber/Benthin, in publication) which should be multi-layered, procedural and temporal. Design criteria for network evaluation should be a multitude of perspectives, process-, future- and identity-orientation as well as an orientation toward a multi-layered approach.

Depending on a given context of economic sectors or institutions, instruments of quality management can be employed, or self-evaluation or ex-post evaluation by experts can be seen as practical. Network monitoring and evaluation which are geared to future-oriented learning and collective development of competence will be designed in a rather decentralized, dynamic and open fashion, although the employment of quantitative methods is not excluded. Evaluative learning arrangements combine qualitative and quantitative methods, methods that generate knowledge and those that “measure” success in a methodical mix, and can thus fulfill the different demands of a networking context. To deal adequately with complex relations of cause and effect they should be represented in a complex fashion (Bangel 1999:354).

Guba and Lincoln (1989) suggested an approach of “stakeholder-based evaluation”, one that is participant-oriented and allows the collective definition of criteria and indicators of successful cooperation. Participatory effect monitoring

follows a central evaluative objective of increasing the collective capability to act, of breaking out of old ruts, and of doing things differently and possibly better than in the past (Oels 2003). Its goal is to expand the repertoire of action (Benthin/Baumert 2001), to increase autonomy and to minimize the degree of manipulation and passivity (Oels, in publication). The approach to evaluation, monitoring and planning described as “stakeholder-based evaluation“ is based on a constructivist paradigm and aims at addressing a large variety of perspectives—which can also be contradictory—in order to create a complex picture of the whole. Indicators and criteria for monitoring and evaluation are generated interactively with the actors concerned. Special emphasis is placed on the definition of learning objectives.

On the basis of a participant-oriented approach, instruments of network management can be put to use which can take over planning, monitoring and evaluation functions and in this way fulfill the evaluative functions of understanding, legitimization and optimization. Especially in open, dialogical settings, the objects of evaluation can be regarded as dimensions of social, functional, structural and learning evaluation.

For example, the social dimension in networks can be analyzed with the help of indicators: the Balanced Scorecard is an instrument for the analysis of network functions as objects of evaluation. Structural tensions can be analyzed e.g. with an appreciative evaluation approach, while the dialogical arrangements of Large Group Interventions can provide an evaluative, planning arrangement of network learning. In this way, contexts and procedures of complex (self-) evaluation are created that simultaneously cover the functions of understanding and optimization, and if needed legitimization as well (Ulrich/Wenzel 2004:28).

Large Group Interventions provide strategic agility and risk minimization in fast transformation processes with a high degree of network activity, because they make use of collective intelligence (Königswieser/Keil 2000). As procedures of transformation they follow the systemic paradigm (Bunker/Alban 1997:5). Systemic, open approaches like Large Group Interventions make it possible to regulate the network tensions brought about by system monitoring and evaluation. They create a mode of “pedagogical organizing”, with its quality of experimental practice (Weber 2004, in publication).

A practice that is oriented towards reflexivity and knowledge generation closes the circle of knowledge provided by monitoring, evaluation and planning in the sense of an incremental, spiral-shaped model of evaluation, working with the iterative practice of producing systemic rationality. But it will never produce “complete” results and will always have rational and irrational parts (Windeler 2001:220). This practice of system reflexivity produces a discursive arrangement of ulterior and self-guidance in which a lot escapes the grip of reflexivity, in which unrecognized conditions and unintended results of actions emerge as well as “blind spots”, chain reactions and “reflexively” influenced causal connections. So participant-oriented effect monitoring in networks will always have to try and strike a balance with that which is not known (Kade 2003). For this reason it will escape the myth of technocratic feasibility – and embark on a journey of collective procedural learning.

References

- Atteslander, P. (1976): Sozialwissenschaftliche Aspekte von Raumordnung und Raumplanung. In P. Atteslander (Hrsg.), *Soziologie und Raumplanung*. Berlin u.a.: De Gruyter. pp. 10-71.

- Bangel, Bettina (1999): Evaluierung der Arbeitsmarktpolitik aus Ländersicht. Die Brandenburger Konzeption der adressatenorientierten Evaluation. In: Informationen für die Beratungs- und Vermittlungsdienste der Bundesanstalt für Arbeit. Nr. 45, pp. 3745-3754.
- Benthin, Nicole; BAUMERT, Martina. (2001): Selbstevaluation als Methode der Qualitätsentwicklung, Prozesssteuerung und summativen Evaluation. In: WEBER, S. (Hrsg.): Netzwerkentwicklung in der Jugendberufshilfe. Erfahrungen mit institutioneller Vernetzung im ländlichen Raum. Opladen. pp. 261-280.
- Bunker, Barbara Benedict; Alban; Billie, T. (1997): Large Group Interventions. Engaging the Whole System for Rapid Change. Jossey Bass. San Francisco.
- Castells, Manuel (2000): Elemente einer Theorie der Netzwerkgesellschaft. In: Sozialwissenschaftliche Literaturreisenschau. 2/2000. Heft Nr. 41. pp. 37-54.
- Duschek, Stephan; Rometsch, Markus (2004): Netzwerktypologien im Anwendungsbereich Kompetenzentwicklung. In: QUEM Bulletin Berufliche Kompetenzentwicklung. Heft 3/2004. Berlin. pp. 1-7.
- Gnahn, Dieter (2003): Indikatoren- und Messprobleme bei der Bestimmung der Lernhaltigkeit von Regionen. In: Brödel, Rainer; Bremer, Helmut; Chollet, Anke; Hagemann, Ina-Marie (Hrsg.): Begleitforschung in Lernkulturen. Münster / New York / München / Berlin. Waxmann. Verlag. pp. 92-106.
- Grunow, Dieter (2000): Netzwerkanalyse. Theoretische und empirische Implikationen. In H.-J. Dahme & N. Wohlfahrt (Hrsg.), Netzwerkökonomie im Wohlfahrtsstaat. Wettbewerb im Sozial- und Gesundheitssektor. Berlin: Edition Sigma. Pp. 303-336.

- Guba, Egon G.; Lincoln, Yvonna S. (1989): Fourth Generation Evaluation. London.
- Hanft, Anke (2003): Evaluation und Organisationsentwicklung. Eröffnungsvortrag zur 6. Jahrestagung der Deutschen Gesellschaft für Evaluation e.V. (DeGeVal). 8.-10.11.2003 in Hamburg. EvaNet-Positionen 10/2003. <http://evanet.his.de>
- Hellmer, F.; Friese, Chr., Kollros, H.; Krumbein, W. (1999): Mythos Netzwerke. Regionale Innovationsprozesse zwischen Kontinuität und Wandel. Berlin. Edition Sigma.
- Helsper, Werner; Hörster, Reinhard; Kade, Jochen (2003): Pädagogische Felder im Modernisierungsprozess. Weilerswist.
- Hendrich, Wolfgang (2003): Ansätze interaktiver Evaluationsmethoden in der beruflichen Weiterbildung am Beispiel informell erworbener Kompetenzen. In: Brödel, Rainer; Bremer, Helmut; Chollet, Anke; Hagemann, Ina-Marie (Hrsg.): Begleitforschung in Lernkulturen. Münster / New York / München / Berlin. Waxmann. Verlag. pp.149-161.
- Herrmann, Franz (1998): Jugendhilfeplanung als Balanceakt. Umgang mit Widersprüchen, Konflikten und begrenzter Rationalität. Neuwied.
- Herrmann, Franz (2001): Planungstheorie. In: Otto, Hans-Uwe; Thiersch, Hans (2001): Handbuch der Sozialarbeit, Sozialpädagogik. Neuwied; Kriftel. 2. Auflage. pp. 1375-1382.
- Jutzi, Karin; Wöllert, Katrin (2003): Erfahrungen und Probleme mit Handlungsforschung am Beispiel der wissenschaftlichen Begleitung

- intermediärer Agenturen. In: Brödel, Rainer; Bremer, Helmut; Chollet, Anke; Hagemann, Ina-Marie (Hrsg.): *Begleitforschung in Lernkulturen*. Münster / New York / München / Berlin. Waxmann. Verlag. pp. 129-143.
- Kade, Jochen (2003): *Wissen - Umgang mit Wissen – Nichtwissen. Über die Zukunft pädagogischer Kommunikation*. In: Gogolin, Ingrid; Tippelt, Rudolf (Hrsg.): *Innovation durch Bildung. Beiträge zu 18. Kongress der Deutschen Gesellschaft für Erziehungswissenschaft*. Opladen. pp. 89-108.
- Kappelhoff, Peter (2000a): *Der Netzwerkansatz als konzeptueller Rahmen für eine Theorie interorganisationaler Netzwerke*. In J. Sydow & A. Windeler (Hrsg.), *Steuerung von Netzwerken*. Opladen: Westdeutscher Verlag. pp. 25-57.
- Kappelhoff, Peter (2000b): *Komplexitätstheorie und die Steuerung von Netzwerken*. In: J. Sydow & A. Windeler (Hrsg.), *Steuerung von Netzwerken*. Opladen: Westdeutscher Verlag. pp. 347-389.
- Kirstein, H. (o.J.): *Was ich schon immer über die Balanced Scorecard wissen wollte*. Unter: <http://www.deming.de/efqm/balanscore-1.html> (am 1.10.2003)
- Königswieser, Roswita; Keil, Marion (Hrsg.) (2000): *Das Feuer großer Gruppen. Konzepte, Designs, Praxisbeispiele für Großgruppenveranstaltungen*. Stuttgart: Klett-Cotta.
- Kuhlmann, Stefan (1998): *Politikmoderation. Evaluationsverfahren in der Forschungs- und Technologiepolitik*. Baden Baden. NomosVerlag.

- Lau, Christoph (1975): Theorien gesellschaftlicher Planung. Eine Einführung. Stuttgart, u.a.
- Lynen von Berg, Heinz; Hirsland, Andreas (2004): Zivilgesellschaft und politische Bildung – Zur Evaluation von Programmen und Projekten. In: Uhl, Katrin; Ulrich, Susanne; Wenzel, Florian M. (Hrsg.): Evaluation politischer Bildung. Ist Wirkung messbar? Verlag Bertelsmann Stiftung. Gütersloh. pp. 15-26.
- Messner, Dirk (1995): Die Netzwerkgesellschaft. Wirtschaftliche Entwicklung und internationale Wettbewerbsfähigkeit als Probleme gesellschaftlicher Steuerung. Köln: Weltforum-Verlag.
- Messner, Dirk (1994): Fallstricke und Grenzen der Netzwerksteuerung. In: PROKLA-Zeitschrift für kritische Sozialwissenschaft. Nr. 97. pp. 563-596.
- Oels, Angela (2000): “Let’s get together and feel alright!” Eine kritische Untersuchung von „Agenda 21“-Prozessen in England und Deutschland. In H. Heinelt & E. Mühlich (Hrsg.), Lokale Agenda 21-Prozesse. Erklärungsansätze, Konzepte und Ergebnisse. Reihe ‚Städte und Regionen in Europa‘ Band 7, Opladen: Leske + Budrich. pp. S.182-200.
- Oels, Angela (i.V.a): Großgruppenevaluation in Netzwerken. In S. Weber (Hrsg.), Netzwerklernen. Methoden, Instrumente. Erfolgsfaktoren.
- Oels, Angela (2003): The Power of Visioning. An evaluation of community-based Future Search Conferences in England and Germany. Münster.
- Schäffter, Ortfried (2001): Weiterbildung in der Transformationsgesellschaft. Zur Grundlegung einer Theorie der Institutionalisierung. Hohengehren.

Simmel, Georg (1908): Soziologie. Leipzig: Duncker & Humblot.

Steger, Renate (2003): Netzwerkentwicklung im professionellen Bereich dargestellt am Modellprojekt REGINE und dem Beraternetzwerk zetTeam. Materialien aus dem Institut für empirische Soziologie an der Friedrich-Alexander-Universität Erlangen-Nürnberg. 6/2003. IfeS. ISSN 1618-6540 (Internet). www://ifes.uni-erlangen.de.

Sydow, Jörg (1999): Management von Netzwerkorganisationen. Zum Stand der Forschung. In J. Sydow (Hrsg.), Management von Netzwerkorganisationen. Wiesbaden: Gabler. pp. 279-305.

Sydow, Jörg (2001): Management von Unternehmensnetzwerken – Auf dem Weg zu einer reflexiven Netzwerkentwicklung? In: Howaldt, Jürgen; Kopp, Ralf; Flocken, Peter (Hrsg.): Kooperationsverbünde und regionale Modernisierung. Theorie und Praxis der Netzwerkarbeit. Gabler Verlag. Wiesbaden. pp. 79-102.

Tuckman, B.W. (1965): Developmental Sequences in Small Groups. Psychological Bulletin 63, pp. 384-399.

Uhl, Katrin; Ulrich, Susanne; Wenzel, Florian W. (2004): Einleitung: Evaluation und politische Bildung – was kann man “messen“? In. Uhl, Katrin; Ulrich, Susanne; Wenzel, Florian M. (Hrsg.): Evaluation politischer Bildung. Ist Wirkung messbar? Verlag Bertelsmann Stiftung. Gütersloh. pp. 9-12.

Ulrich, Susanne; Wenzel, Florian M. (2004): Partizipative Evaluation. In: Uhl, Katrin; Ulrich, Susanne; Wenzel, Florian M. (Hrsg.): Evaluation politischer Bildung. Ist Wirkung messbar? Verlag Bertelsmann Stiftung. Gütersloh. pp. 27-48.

- Weber, Susanne (i.V.) (Hrsg.): Netzwerklernen. Instrumente, Methoden, Erfolgsfaktoren.
- Weber, Susanne (2001a). Netzwerkentwicklung in der Jugendberufshilfe. Erfahrungen mit institutioneller Vernetzung im ländlichen Raum. Opladen: Leske + Budrich.
- Weber, Susanne (2001b): Vom „Ist“ zum „Soll“: Partizipative Verfahren und neue Planungsrationalität. In: Keiner, Edwin (Hrsg.): Evaluation (in) der Erziehungswissenschaft. Weinheim und Basel. pp. 255-272.
- Weber, Susanne (2002): Vernetzungsprozesse gestalten. Erfahrungen aus der Beraterpraxis mit Großgruppen und Organisationen. Wiesbaden: Gabler.
- Weber, Susanne (2003): Zur Evaluation von Großgruppenverfahren am Beispiel regionaler Vernetzung. In: Dewe, Bernd; Wiesner, Gisela; Wittpoth, Jürgen (2003): REPORT. Literatur- und Forschungsbericht Weiterbildung. Erwachsenenbildung und Demokratie. Dokumentation der Jahrestagung 2002 der DGfE Sektion Erwachsenenbildung. pp. 110-119.
- Weber, Susanne (2004a): Transformation und Improvisation. Großgruppenverfahren als Technologien des Lernens im Ungewissen. Unveröffentlichte Habilitationsschrift. Philipps-Universität Marburg.
- Weber, Susanne (2004b): Organisationsnetzwerke und pädagogische Temporärorganisation. In: W. Böttcher & E. Terhart (Hrsg.), Organisationstheorie. Ihr Potential für die Analyse und Entwicklung von pädagogischen Feldern. pp. 253-269.

- Weber, Susanne; Benthin, Nicole (i.V.b): Innovation, Wissen, Selbstreflexivität im Netzwerk generieren. In: Weber, Susanne: Netzwerklernen. Methoden, Instrumente, Erfolgsfaktoren.
- Wenzel, Florian M. (2004): Selbstevaluation wertschätzend gestalten – methodisches Vorgehen in 6 Schritten. In: Uhl, Katrin; Ulrich, Susanne; Wenzel, Florian M. (Hrsg.): Evaluation politischer Bildung. Ist Wirkung messbar? Verlag Bertelsmann Stiftung. Gütersloh. pp. 177-196.
- Windeler, Arnold (2001): Unternehmensnetzwerke. Konstitution und Strukturation. Wiesbaden: Westdeutscher Verlag.
- Winkler, Ingo (2002): Steuerung zwischenbetrieblicher Netzwerke – Koordinations- und Integrationsmechanismen. In: Freitag, Matthias, Winkler, Ingo (Hrsg.): Kooperationsentwicklung in zwischenbetrieblichen Netzwerken. Strukturierung, Koordination und Kompetenzen. Würzburg; Boston. pp. 31-55.
- Wolf, Harald (1999): Arbeit und Autonomie. Ein Versuch über Widersprüche und Metamorphosen kapitalistischer Produktion. Münster .Westfälisches Dampfboot.
- Zipp, Gisela (1976): Planungsziele und Planungswirklichkeit. In: Atteslander, Peter (Hrsg.): Soziologie und Raumplanung. Berlin. pp. 72-93.

Client Impropriety

Chris L. S. Coryn⁴¹, Daniela C. Schröter, and Pamela A. Zeller

Requests for proposals (RFPs) often include statements transferring ownership of the content of proposals to the requestor. Thus, evaluators are frequently faced with the problem of responding to a RFP in an unprotected manner, knowing full well that potential funding entities have the legal right to implement these ideas without the submitter's approval. In extreme cases, funding entities have even requested proposals for the purpose of idea-generation only, that is, it was never the intention to fund these submissions, only to use their ideas.

This kind of ethical abuse is neither new nor unique to evaluation. Allowing intellectual property to become the property of the entity requesting the proposal directly influences the evaluator's work and raises several significant ethical issues regarding the contractual statements found in most requests for proposals which give funding agencies property rights to all information and materials submitted to them.

Take the following case, for example. Recently, a request for proposals was issued for an adult drug treatment program. The RFP was of the usual sort; design, expertise and experience, budget, and so on. Two proposals were ultimately selected as the final candidates: the first, a well planned, systematic evaluation with a proposed budget of just under \$100,000; the second a poorly designed effort

⁴¹ Corresponding author: Chris L. S. Coryn, The Evaluation Center, Western Michigan University, 1903 West Michigan Avenue, Kalamazoo, MI 49008, e-mail: christian.coryn@wmich.edu.

budgeted at slightly more than \$10,000. Why the substantial budget differences? The first proposal was submitted by a university-based evaluation unit and the second was submitted by a university professor acting as an independent consultant, local to the city within which the program was based. As such, the second proposal neither included expenses for travel nor the indirect costs associated with university-based research units. Moreover, the independent consultant indicated within his proposal that all work would be conducted by his students as part of a class project and that these students would not be reimbursed for their work. The funding entity decided that \$10,000 was more attractive than \$100,000. Ultimately, the low-cost competitor was funded, but under the premise of utilizing the costlier competitor's plan and design. The following questions arise as a result of the client's decision:

1. Can the costlier competitor's plan and design be comparably implemented by the low-cost competitor at 1/10 of the price? Perhaps costs can be cut dramatically by hiring a local evaluator with access to free labor and university resources, but what the evaluand saves monetarily may be lost in validity and credibility.
2. Does the low-cost competitor have the expertise and competency to implement the costlier competitor's plan and design? It may be reasonable to infer, in some case, that the contracted low-cost competitor has neither the means nor the competencies necessary to effectively implement the competitor's plan and design.

The client may save as a result of funding the low-cost evaluator if the evaluator is able to implement and fulfill the contract as proposed by the high-cost competitor. However, the issues surrounding the contractual statements that allow all submitted

materials to become the property of the funding agency are indeed troubling. Given the current climate of the competitive evaluation market, proposal writers are faced with several poignant questions:

How detailed and precise should evaluation proposals be if they become the intellectual property of the entity requesting them?

Should the funding entity return rejected proposals?

How can we as evaluators protect our intellectual property given that funders have the rights to use all proposals they receive?

If the funding entity uses any or all of the rejected proposals, in full or part, it should—for the sake of integrity—compensate the originator. This could be accomplished in several ways: (i) a fee could be provided for the use of plans and designs, (ii) the evaluator could collaborate for a consulting fee to help execute the evaluation, or (iii) the evaluator could be contracted as a metaevaluator.

The aforementioned example of client/funder impropriety is utterly unacceptable and the repercussions for the evaluation profession are profound. In addition to the impropriety of the client/funder, other relevant ethical concerns are raised. First, the professor discussed in the case example is more than likely not a member of any organized evaluation organization and therefore not accountable to professional standards of conduct, yet he had obviously violated the unwritten standards of conduct expected of a researcher by accepting the contract and using another's work without the consent of the proposal writer. Second, what can be done to alleviate these problems in the future? Some writers of proposals have attempted to take matters into their own hands by explicitly indicating in their proposals that no portion of the submission may be used without their express

consent. Yet if potential clients/funders willingly and knowingly use these materials, unbeknownst to the proposal writer, what can be done? As can be seen, the implications of an epidemic of this kind of client behavior are frightening. It has been suggested here at the Evaluation Center that approaching AEA might be appropriate. We might suggest developing a code of conduct for evaluation clients, and perhaps some defensive strategies such as blacklisting abusers. What do you think?

Managing Extreme Evaluation Anxiety Through Nonverbal Communication

Regina Switalski Schinker

“Many evaluative situations cause people to fear that they will be found to be deficient or inadequate by others...” (Donaldson, Gooler, & Scriven, 2002, p. 261). Donaldson, et al. (2002) use the acronym XEA to describe excessive anxiety and explain that “...there are people who are very upset by, and sometimes rendered virtually dysfunctional by, any prospect of evaluation, or who attack the evaluation without regards to how well conceived it might be” (ibid). A common technique or ‘magic bullet’ to prevent excessive anxiety would not exist in program evaluation.

In his EVAL 600 class (Western Michigan University, October 26, 2004), Scriven stated that evaluators should care about excessive evaluation anxiety for two reasons. First, if XEA can be quelled, getting information from participants should be more fruitful. Thus, evaluators should strive to make evaluatees less fearful of the evaluation process. Secondly, the likelihood of implementing recommendations should be increased if impactees of the evaluation are comfortable with the evaluation process.

The use of communication research may be a unique approach to relieving XEA, one aspect being nonverbal communication. How can evaluators through unspoken messages impact stakeholders?

A gaze broken too soon, a forced smile, a flat voice, an unreturned phone call, a conversation conducted across the barrier of an executive desk—together such nonverbal strands form the fabric of our communicative world, defining our interpersonal relationships,

declaring our personal identities, revealing our emotions, governing the flow of our social encounters, and reinforcing our attempts to influence others (Ebesu & Burgoon, 1996, p. 346).

In an instant, a stakeholder will make an impression regarding the professional evaluator. This impression will be based on a number of nonverbal cues; eye contact, voice pitch and speed, dress, posture, and facial expression. Credibility, trustworthiness, and expertise are often determined through nonverbal communication channels (Ebesu & Burgoon, 1996; Self, 1996).

Eye contact. Averting eyes, shifting eyes, looking at notes for an extended period of time, and blinking excessively all signal untrustworthiness, insecurity, and/or lack of credibility (Burgoon, Coker, & Coker, 1986; Fatt, 1999; Nolen, 1995). If an evaluator greets a stakeholder with a calm, consistent gaze, they are conveying confidence and believability.

Paralanguage. Paralanguage, or vocal cues, include such factors as volume, rate, pitch and pronunciation (Fatt, 1999). At times, vocal cues are more important than words (Nolen, 1995).

Gestures. Gestures like smiling with head nodding, open arms, casually crossed legs, and leaning towards the person of focus (Nolen, 1995) all convey comfort and confidence. Alternately, such gestures as negative head nodding and foot movement in space signify a tense environment (Keiser & Altman, 1976).

Appearance. Nolen (1995) states, that the objective of communication may determine the choice of clothing. For example, if the evaluator wants to imply receptiveness of others, he/she should mimic the dress of those being

evaluated. “Similarity implies receptiveness” (Nolen, 1995). If the evaluator wishes to promote an image of status and expertise (ibid), they should dress more formally than those they are evaluating.

Environmental factors. An environmental factor such as seating arrangement says a lot about the evaluator and their intentions. “A person expecting to exercise leadership typically sits at the head of a table...” (Ebesu & Burgoon, 1996, p. 350). However, close proximity, or face-to-face, communication leads often the most fruitful communication (Burgoon, Buller, Hale, & deTurck, 1984; Fatt, 1999).

In summary, the evaluator can use the above nonverbal cues to exhibit dominance or cultural similarity; closedness or openness. While nonverbal cues occur almost automatically (Palmer & Simmons, 1995), we must try to be cognizant of them. For a more in-depth understanding, it would be wise and worthwhile for every evaluator to take a graduate level course in nonverbal communication to better understand the person and attitude they are portraying and how their communication cues may affect XEA.

My vision for evaluation concerns communicative style. I would like to see evaluators become conscious of the nonverbal messages they are sending to stakeholders and peers. It is through an evaluator’s communicative style that the image of evaluation will be formed. If evaluators are to reduce anxiety and gain a helpful reputation (Donaldson, 2001) they must approach their stakeholders with friendliness, sociability, and ease.

References

- Burgoon, J. K., Buller, D. B., Hale, Jerold L., & deTurck, M. A. (1984). Relational messages associated with nonverbal behaviors [Electronic version]. *Human Communication Research*. 10 (3, Spring), 351-378.
- Burgoon, J. K., Coker, D. A., & Coker, R. A. (1986). Communicative effects of gaze behavior: A test of two contrasting explanations [Electronic version]. *Human Communication Research*. 12 (4, Summer), 495-524.
- Donaldson, S.I. (2001). Overcoming our negative reputation: Evaluation becomes known as a helping profession [Electronic version]. *American Journal of Evaluation*, 22, p. 355-361.
- Donaldson, S.I., Gooler, L.E., & Scriven, M. (2002). Strategies for managing evaluation anxiety: Toward a psychology of program evaluation [Electronic version]. *American Journal of Evaluation*. 23(3), p. 261-272.
- Ebesu, A. S. & Burgoon, J. K. (1996). Nonverbal Communication. In M. B. Salwen & D. W. Stacks (Eds.), *An integrated approach to communication theory and research* (pp. 345-358). Mahwah, NJ: Lawrence Erlbaum Associates.
- Fatt, J. P. T., (1999, June 1). It's not what you say, it's how you say it - nonverbal communication. *Communication World*. Retrieved November 9, 2004 from http://findarticles.com/p/articles/mi_m4422/is_6_16/ai_55580031/print
- Nolen, W. E. (1995, April 1) [Reading people - nonverbal communication in internal auditing](#). *Internal Auditor*. Retrieved November 9, 2004 from http://findarticles.com/p/articles/mi_m4153/is_n2_v52/ai_17003168/print

- Keiser, G. J., & Altman, I. (1976). Relationship of nonverbal behavior to the social penetration process [Electronic version]. *Human Communication Research*. 2 (2, Winter), 147-161.
- Palmer, M. T. & Simmons, K. B. (1995). Communicating intentions through nonverbal behaviors. Conscious and nonconscious encoding of liking [Electronic version]. *Human Communication Research*. 22 (1, September), 128-160.
- Self, C. C. (1996) Credibility. In M. B. Salwen & D. W. Stacks (Eds.), *An integrated approach to communication theory and research* (pp. 345-358). Mahwah, NJ: Lawrence Erlbaum Associates.

Is Cost Analysis Underutilized in Decision Making?

Nadini Persaud

Is cost analysis underutilized in decision making? Research suggests it is. According to several authors, the use of cost analysis is still infrequent. Further, where cost analysis is conducted, it is often poorly done because many evaluators lack the necessary technical skills (Levin & McEwan, 2001).

Some reasons for the underutilization of cost analysis center on difficulties associated with its use. These include: (1) unfamiliarity with the necessary analytical procedures; (2) political or moral controversies in assigning values to input/outcome measures (e.g. determining the appropriate discount rate); (3) determining the extent to which benefits identified and quantified have been caused by the program); (4) determining who incurs the benefits and costs; (5) determining when benefits and costs occur; (6) inability to quantify all costs and benefits; (7) lack of resources to conduct long-term follow up studies; (8) lack of data; (9) data in a form incomprehensible to the evaluator; and (10) difficulties with separating program developmental costs from operating costs (Alkin & Solomon, 1983; Andrieu, 1977; Berk & Rossi, 1990; Fitzpatrick et al., 2004; Rossi et al., 2004; Sewell & Marczak, 1997).

The current underutilization of cost analysis should seriously concern evaluators, policy makers and society at large. Informed decisions require information on both costs and effects. Given that the ultimate societal goal is to optimize the use of scarce resources, cost analysis can play an important role in national planning. The question is “Can anything be done to raise awareness on this issue?” Yes! Leading

evaluation textbooks and journals must take a more active role in promoting cost analysis. In addition, graduate programs and certificate programs in evaluation need to incorporate cost analysis in their course requirements. If evaluators are not exposed to such techniques and trained to use them, they will never be confident they are conducting cost analysis competently

References

- Alkin, M. C., & Solomon, L. C. (1983). *The costs of evaluation*. Beverly Hill, CA: Sage.
- Andrieu, M. (1977). Benefit-cost evaluation. In L. Rutman (Ed.), *Evaluation research methods: A basic guide*. p. 217–232. Thousand Oaks, CA: Sage.
- Berk, A. R., & Rossi, P. H. (1990). *Thinking about program evaluation*. Newbury Park, CA: Sage.
- Fitzpatrick, J. L., Sanders, J. R., & Worthen, B. R. (2004). *Program evaluation: Alternative approaches and practical guidelines*. (3rd Ed.). White Plains, N.Y: Longman Publishers.
- Levin, H. M. & McEwan, P. J. (2001). *Cost-effectiveness analysis: Methods and applications*. (2nd Ed.) Thousand Oaks, CA: Sage.
- Rossi, P. H., Lipsey, M. W., & Freeman, H. E. (2004). *Evaluation: A systematic approach*. (7th Ed). Thousand Oaks, CA: Sage.
- Sewell, M., & Marczak, M. (1997). *Using cost analysis in evaluation*. Tucson, AZ: USDA/CSREES and the University of Arizona. Retrieved on October 15, 2004, from <http://ag.arizona.edu/fcs/cyfernet/cyfar/Costben2.htm>.

Is E-Learning Up to the Mark?

Fundamentals in Evaluating New and Innovative Learning Approaches Involving Information- and Communication-Technology

Oliver Haas⁴²

Abstract

With the introduction of training courses or seminars via the internet it becomes relevant to evaluate and assess these new and innovative forms of learning and teaching. The following article deals with two issues:

- a) Evaluation methods used in web-based learning arrangements that are dependent on standards in evaluation research
- b) The implicit logic of the communication medium “internet”.

The paper concludes by illustrating methods of evaluation and assessment that correspond to both paradigms.

⁴² Oliver Haas (M. SocSc.) conducted university studies in Sociology at the Johann-Wolfgang v. Goethe University of Frankfurt/ Germany and the Free University of Berlin/ Germany. He is currently employed a Technical Advisor by the [German Agency for Technical Cooperation](#) (GTZ) and has worked in Russia, Tanzania, Malaysia, and South Africa. Here he has been involved in Vocational Education and Training projects.

Introduction

Does the internet make us more intelligent? Do we obtain more knowledge, skills and better qualification through its utilization compared to the “traditional” methods of teaching and learning? Does the internet change teaching and learning concepts or even our perception of learning? After the last big technical innovation—print media—had a significant impact on teaching and learning, the most recent learning technologies⁴³ only changed learning strategies in human resource development to a minor extent. However, the most recent development in Information and Communication Technology (ICT)—the world wide web—as a universal medium of exchanging and mediation of information and knowledge has set a new yardstick regarding the accessibility of learners and the dissemination of learning content.

Innovations in general, and innovative teaching and learning procedures specifically, are always under pressure to legitimize their existence and benefits. This becomes even more crucial when their application is highly related to risks and financial expenditure. Efficiency and effectiveness, quality as well as relevance and significance, to name a few, are key criteria under which these procedures are critically assessed. Despite the recent enormous interest generated in learning via the web⁴⁴, the question of acceptance, efficacy and suitability has not been answered fully.

⁴³ E.g. language laboratories, Computer based training, etc.

⁴⁴ The terms “web-based learning”, “online learning”, “internet-based learning”, and “e-learning” are used interchangeably in this paper.

Methodology-driven and comprehensible criteria-based assessments of procedures, events or actions are called “evaluations”. So far evaluations of internet-based learning have been mainly applied in the public sector. Yet over the past years, the question of efficacy and efficiency of online-learning has become increasingly relevant for the private sector. With regard to in-house and on-the-job training the common understanding is that any training is an investment in employees that needs to be justified like any other investment. While the determination of “costs” is relatively easy to define, the determination of “economies of scale” has proven to be a challenge on its own. This is where evaluation becomes relevant.

To evaluate ‘learning via the internet’ requires an accurate preparation of empirical research that can only lead to useful data if meaningful criteria under which the evaluation will take place have been assigned in advance. Up to now evaluation of learning via the web has been a seriously neglected aspect of impact assessment in education and training. As such, theories on how to go about evaluating learning via the net as well as gaining empirical valid data are difficult to find. However, evaluations in this field are not just necessary but also possible.

It is the aim of this paper to provide a selection of possibilities on how to evaluate the efficiency and benefits of online learning. Due to the complexity of e-learning, this is not possible without a methodological introduction.

Components in evaluating online seminars

To evaluate learning via the internet is nothing new. Assessments of learning software (e.g. in the field of Computer Based Training) have been conducted widely and provide useful experiences as well as theoretical concepts, paradigms and procedures. These are methodological aspects that cannot be neglected when evaluating learning via the internet.

In recent years, highly acclaimed work has been done in the development of methodical standards and instruments in evaluating the efficacy of education and training. However, these standards—consisting models, methods, and instruments—can also be utilized in other contexts. Evaluation methods and criteria are often explicitly created for a specific evaluation purpose. This suggests that evaluation—or better evaluation research—should be considered as applied science following a specific need and demand.

Areas and criteria of evaluation

The setting of criteria should be the first step when assessing online training courses. Criteria need to be defined for each and every aspect of the learning scenario. This involves all participants of the training course, the utilized learning material, the pedagogic approach, and technological aspects, as well as guidance in the learning process and technical-administrative support as part of the training course. Criteria are directly related to the quality of online seminars and constitute the foundation of any evaluative approach to online learning. Thus, a clear statement and definition of evaluation criteria is of crucial importance for the whole evaluation process.

Often several areas of evaluation are strongly interlinked and have a significant impact on each other.⁴⁵ However, for a differentiated assessment of online learning courses it is essential to select evaluation criteria for each evaluation area separately. The decision on evaluation areas and its correlating evaluation criteria is to be done at the beginning of the overall evaluation and needs to be specified for each learning program. Yet, there are some criteria that could be defined as

⁴⁵ For example, the influence of the pedagogic approach or the technology on the students' motivation as well as learning success.

typical in evaluating online learning. The following chart provides a selection of these criteria:

Chart 1: Evaluation areas und evaluation criteria for online learning

| Evaluation area | Evaluation criteria |
|---------------------------|---|
| Participants/ students | <ul style="list-style-type: none"> • Acceptance of training course • Drop-out rate • Degree of collaborative learning⁴⁶ • Rate and intensity of interaction with learning content • Learning success • Communication among students • Transfer and utilization of learning content at the workplace |
| Pedagogical approach | <ul style="list-style-type: none"> • Learning and teaching methods • “Didactic of activation”⁴⁷ • “Didactic of enabling”⁴⁸ • Degree of “blend” in the pedagogic approach⁴⁹ |
| Learning material | <ul style="list-style-type: none"> • Editing and processing of learning content • Comprehensibility, amplitude, correctness, “time sensitivity” of learning material |

⁴⁶ User of training courses can collaboratively work on tasks independent of time and space.

⁴⁷ For a detailed explanation of the term please refer to page 4.

⁴⁸ see above.

⁴⁹ “Blended learning” is an integrated learning concept that combines Information and Communication Technology (ICT) with “traditional” learning methods and media in a single learning arrangement.

| | |
|----------------------------|---|
| Technical system | <ul style="list-style-type: none"> • Quality and reliability of connectivity • Technical infrastructure at the learning site (e.g. internet accessibility) • Collaboration and communication tools |
| Support and administration | <ul style="list-style-type: none"> • Registration and financing • Online-support • Offline-support • Technical support |

Participants

When evaluating online training courses it is crucial to remember that it is not the learner that is evaluated but the learning content delivery! However, the learner plays an important role as a resource person for evaluating the overall training program. The learner’s behavior, learning success as well as transfer of learning content to the workplace provide important empirical information with regards to the quality of the training course.

Pedagogic approach

Even though internet-based learning has been—especially in its early years—strongly related to technology, it is still about the provision of learning content and qualifying people in order to create employability. Therefore web-based learning also—or even more—has to be based on a pedagogic foundation, generally provided through the Curriculum. In Education and Training one distinguishes between two didactic models:

*“Didactic of activation”*⁵⁰

Coming from engineering science, the “didactic of activation” assumes that successful learning can only take place if it is adequately planned with learning methods having been selected accordingly and the sequence of learning is being followed rigidly. Programmatic learning and curricular planning are core determinants of this model.⁵¹

*“Didactic of enabling”*⁵²

This model focuses on the learning and its success. It tries to create an enabling environment for the learner to build up on existing knowledge and to expand his skills and competencies according to his need and demand. Group work, the provision of several learning paths and methods to acquire knowledge are key features of the “didactic of enabling”.⁵³

Degree of “blend” in the pedagogic approach

The “blended” learning approach treats web-based learning as one way of delivering learning content. Consequently, other methods of “traditional” learning

⁵⁰ German: “*Erzeugungsdidaktik*”

⁵¹ See Arnold, Rolf/ Schüssler, Ingeborg, 1998

⁵² German: “*Ermöglichungsdidaktik*”

⁵³ The “action-oriented learning approach” is one relevant learning approach of this model. The action-oriented learning approach is based on a holistic interpretation of technical, individual, methodological and social competence. Learners graduating through this approach are expected to have acquired not only skills and knowledge obtained from qualifications, but also “key competencies”, such as problem solving techniques, communication skills and the ability to work in teams (see Heitmann, Werner, 2004).

such as group or individual exercises remain valid and relevant. Depending on the Curriculum, it has to be decided which learning content will be delivered via the web and where other forms of learning material are more relevant. The demand for content delivery via the web must result from the pedagogic approach of the overall learning program.

Learning material

The learning material provided should be a main focus of the assessment in two respects:

- 1) Quality of content
- 2) The processing of learning content into learning material incorporating ICT

In terms of quality of content it is mainly comprehensibility, coverage, correctness and “time sensitivity” of learning material that needs to be evaluated. Concerning its processing of content into learning material the evaluation should focus on the conversion of content to learning material (e.g. the utilization of text, pictures, animation, simulation, etc.).

Technical System

When dealing with web-based learning, technical aspects are of great importance. Quality and reliability of connectivity including time needed for loading frames and content need to be assessed. The stability of the Web-server as well and its capacity are aspects that must also be assessed. Additionally, the extent and time of utilization of web-content and services provided by the learning platform can provide useful information on the suitability of the technology.

Support and administration

E-learning can take place on an independent level or as an add-on to teaching (the 'blended' approach). The content-based support provided by tutors and technical administration ensures a smooth operation of the e-learning course. The more complex a learning course the more support it needs. However, support and administration can be reduced if training courses include collaboration tools, where students can interact and exchange views, ideas and information. Newsgroups and mailing lists are a very economic way of using portals of this kind.

Types of evaluation

Evaluation research deals with several different types of evaluation. Each has its own suitability to assess the quality of projects and programs and make recommendations for improvement. In the following sections, various types of evaluation will be outlined. However, the focus will be on the suitability of evaluating learning arrangements in the field of web-based learning.

Formative and summative evaluations

Formative evaluations focus on the training course during its development. These types of evaluation aim at ensuring quality and provide useful suggestions for further improvement and refinement of the online course. Summative evaluations however examine the training course after it has been finalized. Here, the main focus is on data compilation. These data give useful hints regarding acceptance, effectiveness and the impact of online training courses.⁵⁴

⁵⁴ See Kromrey, 1998, p.100

Product-evaluations and process-evaluations

Product-evaluations consider one specific product when assessing (e.g. a learning program). In contrast, process-evaluations focus on procedures, handling and utilization of these products.

Internal evaluation and external evaluation

Those who have been actively involved in the development of the online-course do internal evaluation. They conduct the assessment and evaluation of its performance. External evaluation involves external assessors as a main resource to conduct the evaluation.

These types of evaluation provide a grid that should support the decision-making process, showing what type of evaluation is most suitable for which specific training course. Here, suitability depends very much on the type of training course, the state of its implementation, its composition, the assessor's perspective and most importantly the reasons for conducting the evaluation. Obviously a "one-size-fits-all" solution is not possible.

Evaluation methods and gathering of data

Evaluation criteria and empirical data are two central elements of every evaluation. Without evaluation criteria, the acquisition of data can quickly turn into a wild collection of data without correlations, interaction and structure. On the other hand one can say that without any empirical data, questions and presumptions remain without an answer.

In an evaluation, empirical data provide the foundation to validly answer theory-driven questions, to get clarity on assumptions and hypotheses and to be able to

make recommendations. That is why when evaluating online training courses, methods of empirical research are necessary in order to help collect data without interfering with the actual learning process. Generally, internet-technology provides a number of possibilities to collect empirical data. This is especially relevant for written assessment methods (e.g. questionnaires, rating scales, etc.) that are applied frequently in evaluations.

Like all methods of empirical research, evaluation research can be divided into “reactive” and “non-reactive” procedures. When conducting “reactive” assessment procedures (such as interviews) the interviewee is aware of the assessment being conducted. Therefore the person assessed can react to the assessment process in an unpredictable way. The answers can mingle with the originally intended aim of the assessment in such way that it is difficult to make a distinction of all results obtained afterwards. “Non-reactive” assessment procedures (such as hidden observations) take place without the awareness of the subject of assessment. The reactive element of responding does not exist and therefore also no distinction of data is necessary after the assessment has taken place. As a consequence one can say: The assessment with reactive assessment procedures causes a wild mix of (interesting and not interesting) data.

Nevertheless, the utilization of non-reactive assessment procedures also has its challenges: After the assessment is finalized the subject of assessment should be informed on the objectives of the evaluation as well as the reasons for conducting the assessment in such a way. Generally, anonymity of all gathered data in reports should be guaranteed.

Therefore, when selecting evaluation methods it must be noted that obtaining empirical information depends on the decision whether to utilize reactive or non-reactive methods of assessment. A substitution is not possible.

Analysis of documents

Text and documents (e.g. curricular text, teaching text, documentation of communication amongst the participants including the mentor, etc.) on various levels play a crucial role within learning arrangements involving ICT. It is self-evident that all documents used in the learning context need to be carefully tested. When dealing with online training courses this issue becomes even more relevant and higher standards need to be set. This is because the mentor, who functions as a corrective element in the learning process, is not always directly available. It is advisable to make use of programmes designed for text analysis to analyse comprehensive text.

Interviews

Interviews are the most common method of data collection. Here, various variations of data assessment exist. The most popular distinction is made between oral interviews and written interviews. Oral interviews are mostly distinguished by their degree of complexity and structure.

- “Structured interviews” are based on a guideline that has been prepared beforehand. This guideline contains all relevant, necessary and already formulated questions to be asked during the interview as well as (if necessary) hints and tips with regard to the behaviour of the interviewer.

- “Semi-structured interviews” are based on clusters and groups of questions and topics to be dealt with during the interview. A specific sequence or wording of questions is not part of the guideline.
- “Freely structured interviews” are completely free of structure.

Written interviews or questionnaires can be divided based on the form of question, which is used in the interview.

- “Open questions” provide the interviewee with the possibility to formulate the answer in an individual manner.
- “Closed questions” suggest answering options to be chosen from.

Online-questionnaires should be the most common way of assessing and evaluating online learning processes.

Observation

Observations do not play a significant role when assessing online learning. Behaviour can be validly recorded through technology-based recording of behaviour via the usage of the learning platform.

Recording of behaviour

Behavioural expressions in the context of e-learning can be recorded via the so-called “log files” to be found through the server, where all html-documents of an online course are stored. These are access data that provide useful information for the evaluation process (e.g. acceptance of specific learning content, etc.). Besides that, “log files” inform about time, sequence and duration of utilization. However, one should not overestimate the role and function of these files and the information

provided. On the one hand they do not comprise all of the information necessary to conduct an evaluation.⁵⁵ On the other hand they only provide clustered quantitative information that only might be of limited value for the overall assessment. After all, “log-files” only cover the access to HTML-documents. If other forms of communication like collaboration tools, bulletin boards are being used as part of the learning process it will not be captured by these files.

Testing

In evaluation research the term “testing” describes a standardized procedure to assess the occurrence of empirically defined performance characteristics. Usually assessment via “testing” is done in an ad hoc way and rather informally. Standardized forms of testing however are highly sophisticated and complex (e.g. intelligence tests). In standardized testing one distinguishes between “norm-oriented” and “criterion-oriented” methods.

- “Norm-oriented” methods assess an individual test result compared to a control group.
- “Criterion-oriented” methods are based on a predefined figure to assess individual test results.

Empirical research

When assessing online learning, methods of empirical social research need to be adopted and adjusted according to the specific need and demand created by the training course. This means that the whole range of research methods is of relevance for assessing online learning. Just as in any other area of empirical

⁵⁵ If passwords have been given out, it will not be captured by “log files”.

research the decision on the most suitable method depends on the research question as well as the capacity of the respective research method.

Concepts of evaluation for web-based learning

So far we have dealt with evaluation of web-based learning from a methodical point of view. The following sections will apply the outlined methods with the aim of illustrating three types of evaluation concepts. Each one has aspects relating to data survey and data assessment. In other words, after dealing with evaluation of online-learning on an operative level, the following section will provide a “bird’s eye view” of the topic. The first two sections will illustrate aspects of data survey with regards to concepts of evaluation for web-based learning

Utilization of criteria indices

In the field of learning software, “criteria indices” are widespread. Generally, criteria indices can be described as checklists. Evaluation via criteria indices is based on a selection of various relevant and non-relevant criteria that have been pre-determined by experts. Due to their low-cost implications and easy application throughout the whole evaluation phase as well as transparency, criteria indices are very popular. Methods like these are relevant to obtain prompt results and to get a preliminary orientation for the overall implementation process. Furthermore, results gained through criteria indices are easy for others to understand.

However, they also have risks and challenges. These indices are often not based on sound theoretical ground. This shows in uncertainties when selecting criteria as well as the emphasis on each and every criterion. Furthermore, it has been empirically proven that assessors who utilize criteria indices when assessing

learning programmes obtain results that may deviate relatively strongly from each other.

Despite all criticism, one should not generally object to the method of criteria indices. They provide a grid for the area of web-based learning and a possibility for empirical research and evaluation. When expanding these indices through variables (e.g. drop-out rate), one can gain a proper instrument, with which online-seminars and web-based learning programmes can be validly evaluated. To extend this approach with research on students (e.g., individual self-assessment on learning progress) provides further possibilities of application that will be illustrated in the following section.

Determination of coherence

A step beyond the application of criteria indices are those concepts of evaluation, that not only deal with the existence or non-existence of characteristics, but also touch on the elaboration of coherence amongst characteristics. However, this is not free from difficulties. If coherence has been determined (e.g., between the amount of participating students and contributions in online discussions), or differences (e.g., difference between achieving a learning objective and learning groups), the finding is almost impossible to predict for other online-learning programmes if no other variables such as motivation of students are rigidly controlled.

Just as with criteria indices it might also make sense to analyze coherence relations via surveys involving training participants. In fact this is a requirement for many criteria and cannot be replaced by expert surveys.

Evaluations via criteria indices make it possible to allow statements on the theoretical impact of e-learning. A continuation of this approach leads to so-called

“linear structure equalisation models”, where a specific variable (e.g. estimated success in learning) will be determined from interrelations with other variables.

Chart 2: Concepts of evaluation

| | Evaluation via criteria indices | Evaluation via analyzing relations |
|----------------------------------|---|---|
| Perspective of the expert | <ul style="list-style-type: none"> • Learning success • Drop-out rate | <ul style="list-style-type: none"> • Relation between learning success and acceptance of a specific learning/ teaching method |
| Perspective of the user | <ul style="list-style-type: none"> • Self-estimated learning success • Self-estimated degree of communication and interaction with other participants | <ul style="list-style-type: none"> • Estimated relation between the degree of communication with other participants and own learning success |

Aspects of assessing data

If the data assessment is based on criteria indices or interrelations, data alone only partially gives a statement on the quality of online-learning (including suggestions for improvement). The following section will provide an insight into various types of data that can be obtained when evaluating e-learning:

Data indicating learning success

A quantification of learning success only becomes a valid empirical statement through comparison (e.g. before/after-comparison) or through inclusion of analyzing relations (e.g. connection between learning success and participation in a specific learning module/ qualification module).

Decisions of participants/ experts

Judgments made by participants or experts are a fundamental empirical finding when evaluating online-learning. Before making changes based on such statements the data need to be examined further in order to ensure reliability and interrelations with other data.

Data relating to technical features

As long as these data are not utilized in relation with other relevant data of the evaluation (e.g. drop-out rate, learning success), these data can be considered as insignificant.

Data relating to the acceptance of the offered learning programme

Except for the case when responses in this regard are not accessible data, of this kind proves to be very difficult to interpret. Assuming that out of several alternative learning modules only one is accepted by a single learner or small group, it can still be of significance and relevance for a specific learning gproject.

Conclusion and discussion

Despite remarkable work being done in related fields such as educational software, evaluation research in ICT is still very much in its initial steps. However, one can assess online-learning on a practical level as long as an adequate system of classification comprising all relevant and necessary evaluation aspects has been developed. The system presented in the present paper is in conformity with these functional requirements as it provides a pragmatic, criteria-based evaluation that focuses on interrelations and thus serves the purpose of verifying or falsifying hypothesis-based evaluations.

References

- Arnold, Rolf; Schüssler, Ingeborg (1998): *Ermöglichungsdidaktik. Erwachsenenpädagogische Grundlagen und Erfahrungen*, Schneider Verlag, Baltimore.
- Heitmann, Werner (2004): The action-oriented learning approach for promoting occupational performance and employability. South African-German Development Co-operation, Skills Development Research Series, Book 3, Pretoria.
- Kromrey, Helmut (1998): *Empirische Sozialforschung*, 8th edition. Leske + Budrich, Opladen.

The Problem of Free Will in Program Evaluation

Michael Scriven

A group of hard-nosed scientists who have been studying the major commercial weight-loss programs recently reported their disappointment that the proprietors of these programs refuse to release data on attrition. The evaluators, though that's not the label they use, think it's obvious that this is a—or perhaps *the*—key ratio needed to appraise the programs, and one that the FDA should require them to release. On this issue (possibly for the first time in my life), I find myself taking sides with the vendor against the would-be consumer advocate, and I think the issue has extremely general applicability. My take is that the key issue is whether the program, *if followed*, will produce the claimed results; and that following the program is (largely but not entirely) a matter of strength of will. Failure to stay with the program—that is, attrition—is therefore (largely but not entirely) a failure on the part of the subject not the program, and the program should not be 'charged' with it.

First, here's why I think this is a very general problem that we need to deal with, in evaluation overall, not only in program evaluation. Think about the evaluation of: any chemical drug abuse program; twelve step programs like AA for alcohol and gambling abuse; distance or online education; continuing education of any kind—this clearly applies to all of them. Now it also applies in some important cases outside program evaluation, ones that you might not think of immediately. Here are two: (i) it applies to standard pharmaceutical drug evaluation because there is a serious problem referred to as the fidelity or adherence problem, about the extent

to which patients ex-hospital do in fact take the prescribed dosage on a regular basis. In these studies we surely want to say that the merit of the drug lies in what it does if it's used, not whether it's used. Case (ii): in teacher evaluation, although we want to say that the teacher has some obligation to inspire interest, to motivate, as well as to teach good content well, success is clearly limited, not only by natural capacity—as we all agree—but also by dogged disinterest. We don't want to blame teachers for failing to teach inherently capable students who are determinedly recalcitrant, i.e., for high failure ('attrition') rates where the cause is simply refusal to try.

Here's the schema I recommend for dealing with this kind of consideration. Think of a program (or drug regimen, or educational effort) as having three aspects that we need to consider in the evaluation: (A) Attractive power; (B) Supportive power; (C) Transformative power. For short: Appeal, Grip, and Impact. A is affected by presentation, marketing and perhaps allocation, and controlled by selection. The vendor or provider has the responsibility to use selection to weed out cases who are demonstrably unsuitable for the treatment; but, given the unreliability of such selection tests in the personnel area (pharmacogenomics is the subject devoted to this in the pharmaceutical area, where it's considerably more successful) and the importance of giving people a chance when they want to try, one can't be very critical of high-pass filtration for weight-loss, distance ed, and twelve-step programs. Of course, high front-end loading of payments may be excessive, if there's no money-back guarantee.

B is affected by support level including infrastructure (e.g., equipment, air conditioning, counseling), continuing costs (including opportunity costs and fees), and ease of use, for all of which the program is largely responsible; but of course B is also controlled by strength of will. If the support, costs, and ease of use are

disclosed in advance and are both reasonable and delivered as pictured and promised, willpower becomes the controlling variable. Which leaves C, the Impact issue, the real kick in the program: will it deliver as promised if we do our part, taking the pill, doing the homework, getting to the meetings? That's the key issue. While the good evaluator absolutely must check to see if the provider has indeed provided what was promised, and that what was provided was about as good as can be provided at the cost level in question, the rest is up to the subjects. Under these conditions, easily checked and often met, attrition is your failure, not the vendor's.

This is an important issue because it's important that evaluation not assume that these treatments are done to people, and are at fault if they don't work. The fact is that they are selected by people as something *they will undertake, not undergo*, and failure is often the fault of the people not the program. Even with drug treatments, the drugs have to be taken, and often taken for the rest of your life. They only work if you make them work. This is not surgery, which you *do* undergo, which *is* done to you; it's something where you choose to get some help in doing something to yourself. You have to take responsibility for doing your part, and the evaluator must not take that responsibility away and say that the program failed if it didn't get you through to the Promised Land, when it was you who failed. We have free will, but that doesn't mean *success* is a free lunch. Free will is the freedom to start a program: will power is what it takes to complete it.