The political economy of aid evaluation

Michaelowa, Katharina ; Borrmann, Axel

The Political Economy of Aid Evaluation: The German Case

Katharina Michaelowa and *Axel Borrmann*, HWWA[1]

in: Nord-Süd Aktuell, Vol.XVIII, 2004, No.1, pp. 147-155

A. The issue

Today, official development assistance (ODA) exceeds 50 bn € p.a. Given the tight overall budget situation in many of the donor countries, taxpayers can expect the ministries in charge, as well as their implementing agencies, to evaluate the efficiency, effectiveness and sustainability of their activities in a credible way and to report to the public respectively. While in former decades of development assistance, donor agencies provided only limited insight into the reality of aid, there have been increasing efforts for better accountability in more recent years. However, improved evaluation and more frequent reporting could not resolve all doubts in the public. In particular, high degrees of aid effectiveness published in nicely designed brochures have been less convincing than expected by donor agencies. On the contrary, they have inculcated scepticism and provoked questions as to the objectivity of the evaluation findings.

The intention of this article is not to reassess previous studies of donor agencies on the effectiveness of their aid activities. We rather concentrate on their evaluation procedures and the credibility of evaluation outcomes. In this context, we shall attempt to answer the following questions:

- Who has a genuine interest in evaluations and why?
- To what extent does this interest influence the way evaluations are carried out (focus, methodology, selection of evaluator)?
- Which evaluation outcomes are to be expected under these conditions?

These issues will be addressed using a political-economic approach which will also enable us to derivie institutional settings conducive to realistic evaluation results and evaluation-based learning. In the framework of the *Theory of Agency* and the *Theory of Bureaucracy* it has frequently been shown that monitoring and evaluation are actually reliable instruments to

---

1 This paper presents the main outcomes of our more theoretically oriented study " Wer evaluiert was, wie und warum? - Eine politökonomische Analyse am Beispiel der deutschen Entwicklungszusammenarbeit", which will be published in: Ahrens, Heinz (ed.): Zur Bewertung der Entwicklungszusammenarbeit, Schriften des Vereins für Socialpolitik, Berlin. We are indebted to the evaluation departments of BMZ, GTZ and KfW for providing the evaluation data used in our analysis.

solve the information problem in *principal-agent-relations*.[2] From this theoretical point of view, the ultimate objective of an evaluation is to control the principal's agent. Evaluations should serve as an incentive for the agent to perform a task assigned to him so as to fully meet the principal's interest.

In reality, however, evaluations often have a multiple function. They do not only serve as a control for the agents, but are also expected to promote individual and institutional learning. Moreover, they are often intended to strengthen the principal's own accountability with regard to the projects and programmes he is responsible for. If, however, the principal has a preference for specific (optimistic) evaluation outcomes, this can be expected to have a negative bearing on the objectivity of the evaluators and their reports.

There is only a small body of literature addressing this evaluation dilemma in the field of development co-operation. A study by *Mummert et.al* (2002) is probably the only formal political-economic analysis of this problem. In addition, *Carlsson*, *Köhlin* and *Ekbom* (1994) as well as *Nitsch* (2003) feed the discussion by emphasising the advantages of the political-economic methodology. *Easterly's* (2002) fundamental critique of current donor practices also suggests this type of approach. And finally, Brüne (1998) and Kadura (1995) provide some related practical insights. We draw from these studies, but try to extend the theoretical framework to formalise the interrelations and interactions of various actors in the evaluation process in a political-economic model. The model allows for the formulation of a number of hypotheses which will be discussed in the light of empirical evidence for Germany. Finally, some recommendations are derived from the theoretical and empirical analysis, which may contribute to an improved credibility and usefulness of future evaluations in the field of development assistance.

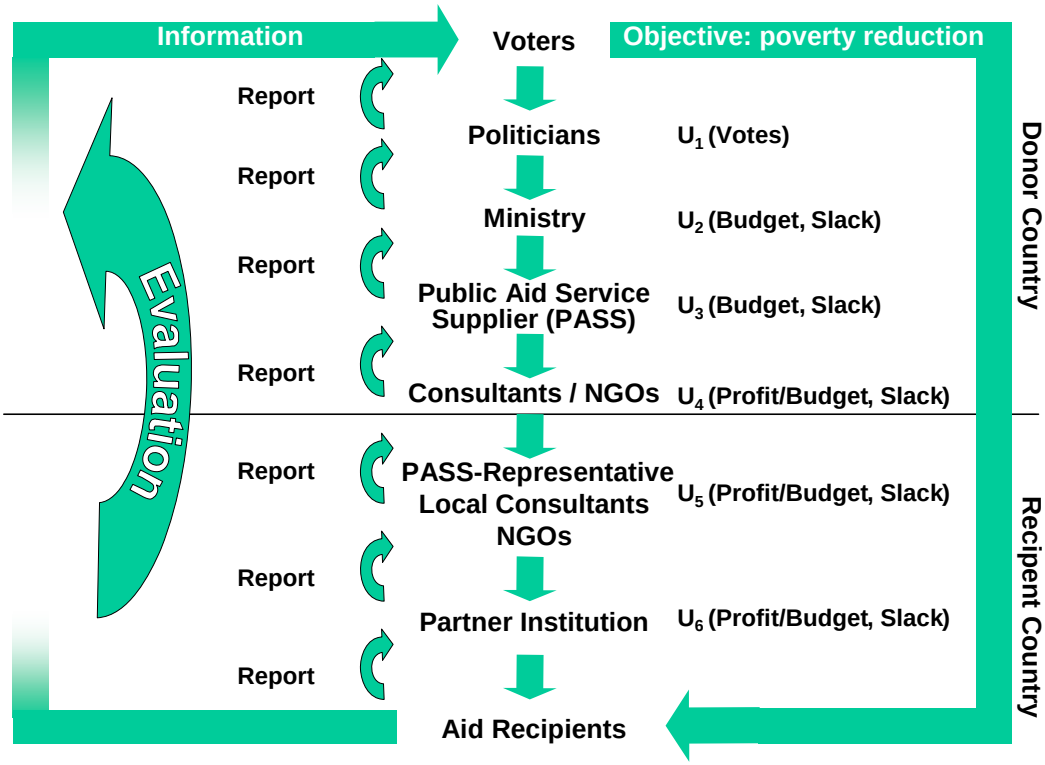
B. Evaluation in development co-operation : some theoretical insights

Development co-operation differs from other policy areas because taxpayers of industrialised countries, who give policy makers a mandate for the delivery of aid, do not benefit directly from the results achieved. Mostly, there is a considerable geographical, political, social and cultural dislocation between the direct beneficiaries of recipient and the citizens of donor countries. Various intermediate institutions are involved in supplying aid services. Taxpayers of donor countries normally have no possibility to get in touch with aid recipients. And foreign beneficiaries have no voting rights in donor countries and thus no political leverage on

2 See: Dixit (2000); Wintrobe (1997, pp 439); Townsend (1979); Mookherjee and P'ng (1989); Moe (1997, pp 464); Melumad and Mookherjee (1989).

donor politicians. As a result, there is a *broken feedback loop* that induces a performance bias in aid programmes (Martens et.al. 2002 pp 154-155). Graph 1 depicts the constellation of players, where voters mandate politicians to provide official development assistance (for example to alleviate poverty), and this mandate is delegated to the government or the relevant ministry and further on to public and non-governmental aid service suppliers (covering non-profit organisations and private consultings). In partner countries, the mandate is forwarded to representatives of donor aid agencies, local consultants and NGOs as well as to partner institutions. It would be possible to include additional layers and their actors within these levels and institutions. However, this would further increase the complexity of our model.

Graph 1: The broken feedback-loop



From this perspective development co-operation appears as a ***hierarchical system of principal-agent relations***. Members or institutions of each level aim at their specific objectives, which do not necessarily coincide with the objectives of their direct principal or the mandated objective of voters, here: poverty reduction. Their objectives can be expressed by a set of stylised utility functions, assuming that the utility depends upon public approval (votes) in the case of politicians, upon their budget in the case of public aid service suppliers, upon profits in the case of private suppliers and, in the case of all suppliers upon pleasant working conditions (slack). Each report requested by the principal from his direct agent passes the filter of the agent's specific utility function. In such a framework, it appears hardly possible that citizens sitting at the end of the pipe are supplied with somewhat realistic information on the effectiveness of aid. However, taxpayers depend on this filtered

information because it would be far too costly for them to obtain first-hand insights on their own. In order to bridge the broken feedback loop and compensate for the information cost disadvantage of taxpayers, evaluation seems to be an appropriate mechanism (*Martens* et.al. 2002, p 155).

Evaluation implies costs that the principal has to bear; however, it allows him to control the services of his agent(s) and actually provides the agent with the required incentive to perform in the principal's interest. Moreover, true reporting can be stimulated by sanctions to be imposed in case an evaluation reveals false declarations by the agent (see for example *Mookherjee* and *P'ng* 1989). This clearly explains that evaluations are of particular importance in the field of development co-operation or in any other policy field where direct feed-back channels - like direct information from programme users - are not available (*Martens* 2002, S. 155).

However, as mentioned earlier, evaluations are not merely used as a control for the agent but also, among other things, to demonstrate the effectiveness of the development assistance as requested by the citizens. In fact as each principal up to the ministry is himself the agent of some other principal up to the voter, each of them has a genuine interest to let his work appear successful. If evaluation results are used for accountancy purposes, for each principal, there is a trade-off between requesting truthful evaluations which will imply the strongest working incentives for the agent, and positively biased evaluations which will make the principal himself appear very successful. Faced with the expectation of positive results, the evaluator engaged to carry out the project assessment – regardless of whether he is recruited externally or from the principal's own staff – gets into a similar dilemma. If he has to assume a strong preference of the principal for a positive image, he will be inclined to avoid a realistic evaluation in order to please the principal and – in the case of a private, external evaluator – to ensure follow-up-orders. However, in the long run, unduely biased evaluation results may spoil the evaluators reputation so that he faces a trade-off just like the principal.

In addition, the incentives for truthful evaluation results also depend on the evaluator's relationship with the agent. Thinking of the agent at the lowest level of the hierarchical structure as the project manager or "expert" on the ground, it becomes clear that this person can greatly facilitate the evaluator's work through providing him with the relevant contacts and information. Obviously, these contacts may be selective and the information may be biased in order to hide potential problems in project management. But the evaluator may accept that and collude with the agent to the detriment of the principal.

Modelling the individual utility function of each actor involved in the delivery of aid can reveal a clearer picture of the problems resulting from this constellation. In order to control

for the complexity of interrelations to be analysed, the model is reduced to four main actors: (1) The aid agency's expert working in the recipient country; (2) the aid agency as his principal; (3) the evaluator evaluating on behalf of the aid agency and (4) the politicians of the donor country acting as the aid agency's principal and being politically accountable for development assistance on the whole. From our model, which is described in detail in *Michaelowa/Borrmann (2004),* a number of hypothesis' can be derived:

- Hypothesis 1: *Overall success rates as published by development ministries and agencies are typically of little informative value.*
- Hypothesis 2: *The hierarchical level at which the evaluation is commissioned affects the evaluation results.*
- Hypothesis 3: *The evaluator's preference for pleasant working conditions (slack) induces distortions of evaluation results caused by a collusion between evaluator and expert, and leads to an overemphasis of evaluation criteria requiring relatively little effort.*

C. Biased evaluation : empirical evidence for Germany's development co-operation

To find some empirical evidence for these hypotheses for the German case we use data provided by the evaluation units of the *Federal Ministry for Economic Co-operation and Development (BMZ)* and two of its leading executing agencies *Kreditanstalt für Wiederaufbau (KfW)* and *Deutsche Gesellschaft für Technische Zusammenarbeit (GTZ).*[3]

Hypothesis 1: *Overall success rates as published by development ministries and agencies are typically of little informative value.*

Overall success rates, as calculated and published by these organisations, result from an aggregation of ratings of a number of projects evaluated within a certain period of time. Table 1 shows the relative frequency of GTZ and KfW projects, which these agencies classified as successful in recent years.

3 BMZ data are derived from six annual cross sectional analyses' of 278 mid-term evaluations carried out between 1990 and 1996 exclusively by external consultants. KfW data stem from about 1200 ex-post evaluations, prepared by default for all projects between 1988 and 2001, mainly by KfW staff members, only some 30% by external consultants. A sample of these reports were subject to external checks by KfW's auditing company. GTZ data were obtained from a 50:50 mix of completion and mid-term evaluations prepared between 1993 and 2000 amounting to a total of about 1200 reports. These reports are the result of a self-evaluation process by GTZ's operational staff, although external random checks have also been made by GTZ's auditing company. For details of the evaluation system of BMZ, KfW and GTZ see Borrmann et.al. (1999) and Borrmann et.al. (2001), GTZ (2001) and KfW (2002)

Table 1: Share of successful GTZ and KfW projects, 1994-2000

	1994-2000	2000	1998-99	1997	1996	1995	1994
GTZ (final review)	84%	82%	88%	81%	85%	85%	78%
GTZ (project progress preview)	87%	96%	87%	87%	89%	88%	74%
KfW (completion reports)	70%	73%	66%	69%	85%	72%	64%

Technical Notes: Projects are considered „successful" according to the agency's specific marking system. KfW's projects are classified as successful, if the evaluators grade them between 1 to 3 out of 6 when they have to answer to the final question of the overall result. Respectively, GTZ success categories are "very successful", "successful" and "successful with reservations".
Source: GTZ, KfW.

What is the real explanatory power of these results? Has GTZ really been avoiding any mistakes in its current projects since the turn of the millennium? Is KfW really less successful in development co-operation than GTZ? According to these figures, it appears more probable to assume that KfW has been applying tighter evaluation standards. This could likely result from KfW's status as a bank and from the respective perception of its clients from the private sector. They would hardly find success rates close to 100% credible, particularly in the field of development co-operation. However, low success rates of less than 60% would noticeably affect KfW's image. Since there is no objective or exogenous rating scale for total success, it can be calibrated by aid agencies according to a desired range. Overall success rates result from an aggregation of single ratings of a set of different evaluation criteria like effectiveness, efficiency, development impact, sustainability. The agency may include or exclude certain criteria or weigh them discretionarily. Data for BMZ evaluations as shown in Table 2 reveal that average project success differs considerably depending on the criteria selected and the aggregation method applied. In contrast to BMZ and GTZ, KfW's leeway in calculating success is limited, however, since the overall result must not arbitrarily deviate substantially from the rate of return, which is a relatively tight criterion and applied by the bank in ex-post evaluations of its financial assistance projects. Aid agencies providing technical assistance normally are not bound that way.

Table 2: Share of successful BMZ projects according to key evaluation criteria[1]

Indicator	Effectiveness	Sustainability	Efficiency	Overall Success[2]
Share of successful projects	71%	55%	83%	39%
(standard error)	(0.028)	(0.032)	(0.025)	(0.033)

[1]BMZ's projects are classified as successful, if evaluators mark 1 to 3 out of 6
[2]Overall success is defined here as successful according to all three individual criteria simultaneously.

Source: BMZ.

To sum up, evidence found here for arbitrary judgements on overall success supports our first hypothesis. Therefore, it seems that current evaluations are actually little effective as an instrument to bridge the broken feed-back loop mentioned above. However, it appears

possible to reduce discretionary power of aid agencies by introducing at least some more objective evaluation criteria such as the rate of return.

Hypothesis 2: The hierarchical level at which the evaluation is commissioned affects the evaluation results.

The underlying idea of this hypothesis is that, at a higher level within the hierarchical principal-agent structure, principals are typically responsible for a higher number of projects so that their own image does not depend so much on the success of any individual project. This should leave more independence to evaluators so that evaluation results can be expected to be less positive. In the German context, we would therefore expect that results of BMZ evaluations tend to be more critical than those produced by GTZ and KfW. Table 3 compares the success rates reported by these different institutions. Unfortunately, data are reasonably comparable only for the effectiveness criterion. They are available for BMZ projects, for a small part of KfW cases and for a fraction of GTZ projects with completion reports. It should be noted that the projects selected by BMZ for complementary evaluations were not selected at random, a fact that may slightly distort results[4].

Table 3: Effectiveness by commissioning agency and type of project

Commissioning Agency	Type of Project	Share of Successful Projects[1]
BMZ	All projects evaluated	71% (0.028)
BMZ	KfW-projects[2]	71% (0.095)
KfW	KfW- projects[2]	74% (0.080)
BMZ	GTZ- projects[2]	67% (0.038)
GTZ	GTZ- projects, final reviews	79%

Technical Note: Standard error in parentheses as far as availability of primary data allows for a calculation.
1 Effectiveness is considered as "successful" if evaluators mark them 1 to 3 out of 6 (KfW, BMZ) or place them in the category "effectiveness > 60%" (GTZ).
2 Joint project of GTZ and KfW excluded.
3 1998-2000 only.
Source: BMZ, GTZ and KfW.

Table 3 reveals that, on average, results for both GTZ and KfW projects are less favourable when evaluations are not commissioned by the respective agency but externally by BMZ. The discrepancy is particularly striking in the GTZ case. KfW ratings differ by 3 percentage points only, a difference that is statistically insignificant. Again, it appears that KfW's self-conception as a bank is also present in its aid activities and seems to be hardly comparable to a mission of a typical aid bureaucracy. Moreover, development assistance plays only a minor role in the banks overall portfolio.

When the commissioning level for evaluations is taken into account, it is no longer justified to consider GTZ projects as more successful than KfW projects as Table 1 might have

4 Some projects were selected because they faced particular problems, or, on the contrary, because they were considered as an example of best practice ("model projects").

suggested. The difference between BMZ ratings of GTZ and KfW projects is insignificant, and in fact, the share of successful projects is actually slightly higher for KfW than for GTZ.

It follows that there is no empirical evidence for hypothesis 2 in the case of KfW, but some support in the case of GTZ. When the comparison between GTZ and BMZ evaluations is extended to additional criteria like planning, implementation, project management and sustainability, even stronger evidence can be found for these findings.

It must be pointed out, however, that beside the hierarchical level of commissioning evaluation, results are also influenced by the degree of institutional independence of evaluators. It is difficult to separate the impact of these two factors. For example, better results produced by GTZ as compared to BMZ ratings of GTZ projects may also be due to the fact that, as a rule, GTZ does not recur to external consultants for final project evaluations. However, a comparison of GTZ project progress reviews and final reviews reveals (see Table 1) that the introduction of external evaluators does not automatically improve the credibility of the evaluation process. On the contrary, GTZ project progress reviews, fully or partly carried out by external consultants, show more favourable overall results than GTZ self-evaluations upon completion normally done by the operational staff. In particular when GTZ entrusted project managers in the field to commission project progress reviews and to recruit the required external consultants, success rates were reported to be at their highest ever and reached a more than dubious level of 96%. Under these conditions, the independence of the so-called “external” evaluator is probably at least as questionable as the independence of an evaluator who is not external to the aid agency, but from a different department not involved in project management.

Hypothesis 3: *The evaluator’s preference for pleasant working conditions (slack) induces distortions of evaluation results caused by a collusion between evaluator and expert, and leads to an overemphasis of evaluation criteria requiring relatively little effort.*

If expert and evaluator coincide, as is the case in GTZ’s final reviews, collusion is at its maximum. But also for normal cases, we expect that the assumed collusion between evaluator and expert leads to a higher-than-average rating of criteria the expert is directly responsible for. This holds particularly true for project implementation. We assume that the evaluator working closely together with an expert in the field day by day and being highly dependent on his support is inclined to attribute project failures more to unfavourable framework conditions than to shortcomings in project implementation, for which the expert is directly responsible. BMZ data in Table 4 actually point in this direction. Project coordination, framework conditions and project planning show significantly lower ratings than project implementation. Efficiency is the only criterion with even more favourable evaluation results. However, this could be explained by the fact that in many cases, efficiency is not assessed at all, so that those projects where it actually is, may be a positive selection of the overall sample.

Table 4: BMZ ratings for key evaluation criteria

Criterion	Share of projects considerd at least satisfactory	Standard error	Number of projects
Implementation	72,0%	0.028	261
Definition of objectives	70,3%	0.027	283
Project coordination	60,7%	0.030	262
Framework conditions	60,6%	0.029	282
Project planning	56,1%	0.030	278
Efficiency	82,7%	0.025	225
Effectiveness	70,6%	0.029	255
Sustainability	54,6%	0.032	249

Source: BMZ.

Missing information on individual evaluation criteria is worth being examined more closely as such. Beside efficiency, missing information is also quite frequent for sustainability. Assessing efficiency and sustainability requires the application of more sophisticated analytical tools as compared to the evaluation of the definition of objectives or of framework conditions, which is more or less a matter of simple description. The number of projects for which these easily obtainable ratings are available is much higher than for sustainability and efficiency. Evaluators often complain that either time and budget constraints or specific project characteristics do not permit a cost-benefit-analysis. Some evaluators circumvent this problem by assessing efficiency and sustainability in a brief, barely founded and intuitive way. Similar behaviour of evaluators can be observed for the development impact criterion which is not included in the BMZ-dataset. It has long been criticised that the Ministry failed to introduce more a sophisticated set of empirical methods to evaluate such a fundamental issue (see for example *Stockmann* 1996). A comprehensive attempt, the first of it's kind, has been made only recently (BMZ 2000).

All in all, there is some empirical evidence for hypothesis 3. To further examine the impact of collusion on the information provided on different evaluation criteria, it would be useful to compare mid-term evaluations and ex-post evaluations where the project manager can no more interfere. This is not feasible, however, since real ex-post evaluations are carried out only by KfW and its dataset does not show separate results for implementation.

D. Starting points for a more efficient evaluation system

Both, from a theoretical and an empirical point of view, the explanatory power as well as the credibility are apparent problems in current evaluations of development co-operation. Critical issues are:

- the evaluator's dependence on the acceptance of the results by the principal,
- the probable collusion of the evaluator and the expert in the field being in charge of project implementation, and
- the limited disposition of the evaluator to introduce powerful, but costly evaluation methods (in terms of time and effort).

To arrive at a more efficient evaluation system, institutional solutions, which change the prevailing incentives of both evaluators and aid agencies, need to be identified.

As to ***greater independence of the evaluators***, GTZ, KfW and other German aid agencies have recently made considerable progress by establishing evaluation units that are distinctly separate from operational line management. BMZ's evaluation function is autonomous even since 1970. Nevertheless, further reforms are required. GTZ still leaves the responsibility for the final project reviews with the local expert in charge of implementation, and GTZ's project progress reviews have lost much of their credibility as an independent evaluation tool since the end of the nineties, when the expert was entrusted with the selection of the evaluator.

In addition, evaluation efficiency can also be improved by strengthening ***external evaluation.*** *Mann (2000)* has shown for cases outside development co-operation that evaluations initiated by institutions that are also in charge of project implementation, have a positive bias. It has been frequently suggested to enhance the independence and professionalism of evaluations by establishing a separate body or appointing a commissioner for development policy outside the system of development co-operation (*Stockmann* 1997; *Stockmann* 1998, pp. 110-113; *Caspari*, *Kevenhörster* and *Stockmann* 2003, p. 14). This should contribute to neutral expertise, to an improvement of evaluation methodology, and to informative assessments of the development impact and the sustainability of projects – criteria which are frequently neglected in current evaluations. However, to exclude additional risks of collusion, an evaluation monopoly should be avoided. Existing institutions like the federal audit court (Bundesrechnungshof), accounting firms, neutral scientific bodies inside and outside Germany as well as the OECD Development Assistance Committee (DAC) should be used instead. Contrary to current practice, these institutions should not limit their activities to spot checks of existing self-evaluations but extend them to full-fledged re-evaluations of a sample of projects. To maintain a high standard of evaluation quality, less frequent evaluations are acceptable. As the principal-agent literature suggests, the incentive for aid agencies to carry out less biased self-evaluations can be improved by a threat of result-based re-evaluations

(with the probability of a re-evaluation increasing if the result of the self-evaluation is positive; *Mookherjee* and *P'ng* 1989, p. 408 f.). Re-evaluations should be used for methodological improvements which could also induce the interest of aid agencies to meet certain minimum standards and to apply state-of-the art tools. For instance, in the framework of the project assessments within the DAC Peer Reviews, the reviewers rather than the agency under review should select the projects to be evaluated. Selection could take place on the basis of prior evaluations carried out by the agency itself so that results of both assessments could be compared.

To reinforce the incentive to meet evaluations standards, for instance those of the DAC (*DAC* 1992) or those of evaluation societies like the DeGEval (*Deutsche Gesellschaft für Evaluation*) (*DeGEval* 2001), evaluations systems of aid agencies should be subject to comprehensive assessments at regular intervals (*Borrmann* et al. 1999, p. 333). Ex-post evaluations including impact assessments should be carried out in greater numbers several years after project completion. This would also avoid the problem of collusion between evaluator and expert.

The incentive structure for achieving better evaluation results could also be strengthened by a higher degree of ***transparency***. Most German aid agencies remain very reluctant in this respect. However, BMZ started a turnaround in 1999 by granting access to its evaluation reports with a few reservations only. Moreover, summaries and BMZ's bi-annual evaluation programme are now available on the internet. Since the early nineties, KfW and GTZ have been publishing their regular cross-sectional evaluation reports. Recently, KfW started to publish a condensed version of each evaluation report including main findings and detailed success ratings (KfW 2002, p. 3). Indeed, this points into the required direction. It mitigates the transaction costs of NGOs and the media in their information retrieval and hence strengthens their function as an external control. Aid agencies run a higher risk of being blamed for all-too positive evaluation results (see also *Martens* 2002, p. 176). If the general public is well informed, it is in the agencies' genuine interest to reduce the evaluation bias and to put less pressure on their evaluators. Improved external access to evaluation data could enhance this effect since independent outside research would be facilitated. This would also stimulate the improvement of evaluation methods.

Finally, the efficiency of the evaluation system could significantly be improved by strengthening market forces and competition, both among executing agencies and evaluators. In practice, the link between success and budget in Germany's development co-operation is rather weak. The allocation of funds to German aid agencies mostly follows political criteria. Current efforts to merge public agencies with similar portfolios will reduce competition even more. To obtain real competition, commissioning should be based on at least EU-wide public bidding, which would give domestic and international private consultants a better chance. The same procedure should be introduced for evaluators. Partiality of evaluators mainly stems

from their experience that positive results increase the prospects for subsequent orders. Therefore, personal contacts as well as "harmonic relations" currently dominate the recruitment of evaluators, whereas professional skills like sectoral know-how, language ability, country and field specific experience play only a secondary role (*Kadura* 1995, p 14, *Brüne* 1998, p. 20). To enlarge the market and to reinforce competition, the recruitment of evaluators should become more transparent in the future. Tendering procedures should not be restricted to a narrow group of consultants highly specialised on advisory services for development assistance; international consultants and those, who have gained experience in other fields, should also be considered. As our model clearly shows, a higher degree of diversification mitigates the evaluator's dependence upon his principal and hence leads to a less biased evaluation result. Diversification is particularly required, if the number of independent principals is very small, which is the case in the field of development co-operation (even if less so for Germany than for most other donor countries). Opening up the market for international consultants would compound this effect. Increased competition would also stimulate efforts to improve evaluation quality in general, and evaluation methodology in particular. Dependence of evaluators on follow-up orders by the same principal would decline.

To give priority to professional qualification rather than to proven opportunism of the evaluator, the recruitment could also be transferred to an institution outside the aid system (*Brüne* 1995, p. 6). Moreover, a maximum period of consultancy for any specific aid agency could be introduced for evaluators in order to ensure their independence. They would then have to work for other institutions during several years before being eligible for the initial aid agency again. While working for any specific aid agency, the evaluator's primary concern would then be his own reputation decisive to obtain follow up orders from other institutions.

Conclusions

All in all, there are a number of promising solutions to improve the efficiency of Germany's evaluation system. Several aid agencies have introduced reforms pointing into the right direction while others lag behind. The whole system still requires increased transparency and enhanced competition. From a political-economic point of view, further progress will primarily depend on the interest of the public. Only if the public shows a stronger interest in realistic information about aid effectiveness, pressure on politicians will increase, and aid institutions will be monitored more closely with respect to the credibility of their results.

References

BMZ (2002): Evaluierungsraster (http://www.bmz.de/themen/erfolgskontrolle/fachinfo_zep/evaluierungsraster.pdf)

BMZ (2000): Langfristige Wirkungen deutscher Entwicklungszusammenarbeit und ihre Erfolgsbedingungen. Eine Ex-post-Evaluierung von 32 abgeschlossenen Projekten, BMZ-Spezial No. 19, Bonn

Borrmann, Axel, Karl Fasbender, Manfred Holthus, Albrecht von Gleich, Bettina Reichl and Rasul Shams (1999): Erfolgskontrolle in der deutschen Entwicklungszusammenarbeit, Analyse, Bewertung, Reformen, Baden-Baden

Borrmann, Axel, Albrecht von Gleich, Manfred Holthus and Rasul Shams (2001): Reform der Erfolgskontrolle in der deutschen Entwicklungszusammenarbeit, Eine Zwischenbilanz, Baden-Baden

Brüne, Stefan (1995): Zwischen Marketing und Erfolgskontrolle: Evaluierung in der entwicklungspolitischen Zusammenarbeit, in: Stefan Brüne (ed.): Erfolgskontrolle in der entwicklungspolitischen Zusammenarbeit, Thesen und Materialien zu einer Tagung des Deutschen Übersee Instituts, Hamburg, pp. 5-6

Brüne, Stefan (1998): Evaluierung als öffentliche Kommunikation - Zu den politischen und entwicklungsbezogenen Rahmenbedingungen entwicklungsbezogener Wirkungsbeobachtung, in: Stefan Brüne (ed.): Erfolgskontrolle in der entwicklungspolitischen Zusammenarbeit, Hamburg, pp. 9-26

Carlsson, Jerker, Gunnar Köhlin and Anders Ekbom (1994): The Political Economy of Evaluation - International Aid Agencies and the Effectiveness of Aid, New York

Caspari, Alexandra, Paul Kevenhörster and Reinhard Stockmann (2003): Das Schweigen des Parlaments – Die vergessene Frage der Nachhaltigkeit deutscher Entwicklungszusammenarbeit, in: Aus Politik und Zeitgeschichte, No. B13-14, 2003, pp. 7-14

DAC (1992): DAC Principles for Effective Aid, Development Assistance Manual, Paris

DeGEval (2001): Standards für Evaluation, Köln,
http://www.degeval.de/standards/Broschuere.pdf

Dixit, Avinash (2000): Incentives and Organizations in the Public Sector: An Interpretive Review, unpublished, Princeton University

Easterly, William (2002): The Cartel of Good Intentions: Bureaucracy versus Markets in Foreign Aid, Center for Global Development, Working Paper No. 4, revised version

GTZ 1994): Erreicht die Technische Zusammenarbeit die gesetzten Ziele?, Projektergebnisse der GTZ und ihrer Partner, Eschborn

GTZ (1996): Erreicht die Technische Zusammenarbeit die gesetzten Ziele?, Projektergebnisse der GTZ und ihrer Partner - Zweite Querschnittsanalyse, Eschborn

GTZ (2001): Bericht zur 7. Querschnittsanalyse, Wirkungsbeobachtung von im Jahr 2000 laufenden und abgeschlossenen TZ-Vorhaben; Stabsstelle 04, Team Interne Evaluierung, Eschborn

Kadura, Bernd (1995): Wie frei ist ein freier Gutachter? Zwischen bürokratischer Gängelung, selbstauferlegter Befangenheit und lasziver Gestaltungsfreiheit, in: Stefan Brüne (ed.): Erfolgskontrolle in der entwicklungspolitischen Zusammenarbeit, Thesen und Materialien zu einer Tagung des Deutschen Übersee Instituts, pp.9-14

KfW (2002): Förderung mit Perspektive: Aus Chancen Zukunft machen. Siebter Evaluierungsbericht über Projekte und Programme in Entwicklungsländern, Frankfurt

Mann, Stefan (2000): The Demand for Evaluation from a Public Choice Perspective, in: Deutsches Institut für Wirtschaftsforschung, Vierteljahresheft No. 3, Evaluation im Spannungsfeld zwischen Wissenschaft und Politik, Berlin, pp. 371-378

Martens, Bertin (2002): The role of evaluation in foreign aid programmes, in: Bertin Martens, Uwe Mummert, Peter Murrell and Paul Seabright: The Institutional Economics of Foreign Aid, Cambridge, pp. 154-177

Melumad, Nahum and Dilip Mookherjee (1989): Delegation as Commitment: The Case of Income Tax Audits, in: Rand Journal of Economics, Vol. 20, No. 2, pp. 139-163

Michaelowa, Katharina and Axel Borrmann (2004): Wer evaluiert was, wie und warum? Eine politökonomische Analyse am Beispiel der deutschen Entwicklungszusammenarbeit; in: Heinz Ahrens (ed.): Zur Bewertung der Entwicklungszusammenarbeit, Schriften des Vereins für Socialpolitik, Berlin, forthcoming

Moe, Terry (1997): The positive theory of public bureaucracy, in: Dennis Mueller: Perspectives on public choice: a handbook. Cambridge, pp. 455-480

Mookherjee, Dilip and Ivan P'ng (1989): Optimal Auditing, Insurance, and Redistribution, in: The Quartely Journal of Economics,Vol. 14, No. 2, 1989, pp. 399-415

Nitsch, Manfred (2003): Evaluationskriterien für Mikrofinanzinstitutionen: Finanzielle Nachhaltigkeit, Erreichung ärmerer Zielgruppen und Einkommenswirkungen;´, Korreferat zum Beitrag von Manfred Zeller; in: Heinz Ahrens (Ed.): Neuere Ansätze der theoretischen und empirischen Entwicklungsforchung, Schriften des Vereins für Socialpolitik, Gesellschaft für Wirtschafts- und Sozialwissenschaften, Neue Folge Band 297, Berlin, pp. 27-38

Stockmann, Reinhard (1996): Die Wirksamkeit der Entwicklungszusammenarbeit. Eine Evaluation der Nachhaltigkeit von Programmen und Projekten der Berufsbildung, Opladen

Stockmann, Reinhard (1997): Stellungnahme zum AwZ-Workshop „Nachhaltigkeit in der EZ", Deutscher Bundestag, Ausschuss für Wirtschaftliche Zusammenarbeit und Entwicklung, Stellungnahmen zu der öffentlichen Anhörung des Ausschusses für wirtschaftliche Zusammenarbeit und Entwicklung, „Nachhaltigkeit in der Entwicklungszusammenarbeit" am 29.10.1997, Ausschussdrucksache 13/162

Stockmann, Reinhard (1998): Viel Kritik - aber wenig profundes Wissen, in: Stefan Brüne (ed.): Erfolgskontrolle in der entwicklungspolitischen Zusammenarbeit, Hamburg, pp. 88-123

Townsend, Robert (1979): Optimal Contracts and Competitive Markets with Costly State Verification, in: Journal of Economic Theory, Vol. 22, pp. 265-293

Wintrobe, Ronald (1997): Modern Bureaucratic Theory, in: Dennis Mueller: Perspectives on Public Choice, a Handbook, Cambridge, pp. 429-454