

56422

**Management and Evaluation within the Plano Plurianual:
Institutionalization without Impact?**

**Yasuhiko Matsuda, Geoffrey Shepherd, and Juliana Wenceslau
November 6, 2006**

Disclaimer

This report is based on field work undertaken in February and May of 2006 by Yasuhiko Matsuda, Geoffrey Shepherd, and Juliana Wenceslau. The report was prepared in response to a specific request from the Ministry of Planning. As it does not constitute an official World Bank report, it has not been subjected to the World Bank's regular review process. The views and positions expressed in the report are the authors' individual views and do not necessarily represent the World Bank's official view on the subject.

1. INTRODUCTION

1. This report was prepared in response to the request from the Government of Brazil to evaluate the PPA's evaluation system and propose ways it could be strengthened. In our understanding, the genesis of the PPA as an instrument of government planning and management and its philosophy was circa 1999. It was then that the decision was made to apply a particular planning model piloted with a set of priority projects (*Brasil em Ação*) in the 1996-99 PPA to the entire set of government activities in the PPA 2000-03. The current state of the PPA reflects a series of evolutionary changes since then, with the improvements in its evaluation framework being one of them. Because of this origin, understanding and evaluating the PPA evaluation system would require placing it in the context of the evolution and the effectiveness of the whole PPA, as envisioned in its original model and its subsequent adaptations. Our point of departure is the observation that the primary challenge is to make the PPA effective as an instrument to make the federal administration more performance oriented. Improvements to the evaluation system as such, though welcome and needed, should be a secondary objective for the government, as this is only a limited part of the whole endeavor.

2. In this report, we present key findings from our rapid assessment, based on a set of interviews with key stakeholders in the PPA process and a review of a small sample of PPA program evaluations. The intent of the report is to flag certain issues as guides for the government's further considerations rather than present definitive findings based on robust empirical evaluations, for which we had neither sufficient time nor resources.

3. As detailed below, our view is that the PPA model itself needs revamping for it to regain its credibility among federal officials and thus its effectiveness as a tool for managing the government's policies and resources with explicit results-based perspectives. Evaluations could play a useful role in making the PPA a stronger instrument of policy and resource management. But evaluations are worthwhile only so long as their results are effectively used for decision-making and actions. In this sense, how the evaluation framework itself can be improved will in turn depend on what aspects of the PPA will be revamped and how. Thus in concluding the report, we offer some options as possible guides for the government's own effort to review and revamp the whole PPA model, and from there derive some implications for the design of the PPA evaluation framework.

2. THE EVOLUTION OF THE PPA

The PPA 2000-2003: *Avança Brasil*

The Model

4. The PPA 2000-2003 combined instruments of planning, budgeting, and management (see World Bank, 2002):

- *Planning.* It revived the Brazilian tradition (from the 1950s to the 1970s) of planning, both through the indicative planning of regional development (*Eixos*) and its attention to public-private and federal-sub-national-government partnerships.
- *Budgeting.* It linked planning (policy priorities) to budgeting (resources) by assigning all expenditures to a program and making the program a principal spending classification in the budget. (But the strict alignment of planned spending under the PPA and budget allocations only held for the first year of the Plan, and even this was accomplished by having the Plan approved by Congress only after the approval of the annual budget.)
- *Management.* It introduced a new management model under which: each program was characterized by specific objectives and indicators and assigned a manager; and the Ministry of Planning's program-by-program oversight of plan execution was supported by a monitoring and evaluation system (supported by a management information system, SIGPlan).

5. The PPA was a unique and ambitious instrument meant to move the federal public administration from traditional ("Weberian") forms of management towards the model of performance-oriented management that has taken hold in many OECD countries.¹ Compared to the other models, which have tended to emphasize strengthening performance incentives and accountability of government organizations, Brazil's version put more emphasis on planning and on organization of government activities by program. The PPA emphasized performance by creating tools appropriate to the programs (program objectives and performance indicators, program managers) and to a performance-oriented budget cycle (annual program evaluations and revisions). But unlike some OECD countries it did not promote specific contractual arrangements (e.g., by hiring managers on explicit performance contracts as in New Zealand), nor did it focus on organizational change (e.g., by providing organizations and units responsible for program execution high degrees of managerial autonomy in exchange for accountability for performance).

6. In many ways, the original conception of the PPA saw the program structure as an alternative to the existing organizational structure of the government. In theory, at least, each of the programs was defined on the basis of a concrete problem in society that it was

¹ The World Bank's report on the PPA and its first year or two of implementation (World Bank, 2002) characterized the reform as cautious, in the sense that proposed managerial changes were flexible and negotiable and did not mandate organizational change. But overall, the reform has proven utopian.

meant to address. Logically, neither the problems in society nor their possible solutions were necessarily confined within the existing organizational boundaries of the government bureaucracy. Thus in its purest form, the PPA programs were conceptualized as supra-organizational units of structuring government actions, resource allocations, and managerial accountability.

7. It can be reasonably inferred that the architects of the PPA 2000-2003 had an implicit agenda of using the program format to counterbalance the power of many ministries and lodge a greater amount of influence over government activities at the center of government (Ministry of Planning). By the same token, this agenda meant to support the technocracy by protecting professional specialists from politics. This was an ambitious agenda, given the multiplicity of political influences on Brazil's federal executive.² The Plan derived some political impetus from the success of its predecessor PPA (1996-1999 *Brasil em Ação*) and built on the management model of that Plan. The PPA 2000-2003 was also a convenient political project for the 1998 reelection campaign of Fernando Henrique Cardoso (Gaetani, 2003). In 1999 it was launched with much political fanfare (e.g., PPA managers appointed in a presidential ceremony). But it did not turn out to be a particular priority of the second FHC Administration as the term progressed, although it is our impression that the whole initiative benefited from the residual momentum even toward the latter years of the FHC2 administration.

Implementation and Results

8. The PPA was implemented pretty much as intended, in the sense that its routines (programs as the basis for the budget and for reporting to the center of government) became the norm throughout the federal administration (i.e., the PPA rapidly became institutionalized at least in form). In substance, of course, the PPA faced a number of challenges as its philosophy implied a radical departure from the usual ways in which government activities were conceptualized and managed.

9. One of the critical weaknesses of the PPA (and Brazil's planning and budgeting system overall), as illustrated in the Bank's previous report on the PPA (and well-recognized within the government), was its difficulty in clarifying government priorities and ensuring financial (let alone physical) execution of those programs designated as government priorities. The difficulty in assuring necessary financial flows to priority programs was due to the extremely high degree of budget rigidity enshrined in the Constitution and other laws that provided protected funding to specific policies and programs. The budget rigidity required sharp definition of priorities so as to direct the

² In spite of the Federal government's longstanding commitment to planning (and evident planning successes during the "miracle" years), it has difficulty in coordinating policy-making. A number of ministries are subject to strong influences from outside the executive, because of the extent to which their workforce corporately exerts influence over them, the extent to which they enjoy protected funding, or the extent to which they have been "colonized" by special interests (often through Congress or sub-national governments). This is not unlike other countries where federal and/or presidential constitutions create divisions of purpose. But the situation seems to be made worse by the fragmentation of policy-coordinating powers, within Brazil's executive, between the presidency, the Ministry of Finance, and the Ministry of Planning (a fragmentation that is mirrored in the lack of a formal high-level forum for policy-making and -coordination – a role played by ministerial cabinets in other countries).

scarce discretionary funds to the small subset of important programs and projects. However, from the beginning the government found it difficult to choose a small enough subset of programs as priorities, and thus was often unable to guarantee funding for all of them.

10. To make the matter worse, opportunistic behavior by line ministries (e.g., so-called priority inversions) further constrained MPOG's ability to enforce execution of the PPA priority programs. To this situation, the most important corrective applied during implementation was the introduction of "cash-flow control" in 2001-02, apparently with some success. But this was a stop-gap measure that could not be applied too broadly and its heavily centralized operation went against the logic of ministerial (or program) accountability for results. Ultimately the solution would only come in the form of improved prioritization within each sector as well as across sectors, and a disciplined approach to budgeting and management whereby the ministries are held accountable for effectively executing their stated priorities.

11. Other difficulties arose from the design of the PPA model, such as the attempt to force inter-sectoral coordination through multi-sectoral programs. This, in particular, was a good idea in theory that met practical difficulty during implementation. Under this model, it was hoped that the managers of the multi-sectoral programs would be able to work across ministerial boundaries without going through the ministries' hierarchical chains of command for day-to-day managerial decisions. As it turned out, however, most PPA program managers were never given the kinds of authorities and responsibilities usually expected of program managers. In many cases, for example, ministry staff who were actually delivering the program did not report to the program manager. Without even the authority to control the entire set of inputs for implementing a given program within their own ministry, these managers were far from empowered to facilitate complex negotiations across ministries to implement multi-sectoral programs.³

12. Our best judgment is that the PPA 2000-2003 constituted a part of a modest move to greater performance orientation that has been going on in the federal government since the first FHC Administration. The PPA promoted performance orientation through several channels. First, most observers agree that the program format provided greater transparency, and in theory at least, greater clarity of purpose through the explicit definition of program objectives and performance indicators. Second, some ministries tended to become more performance-oriented (as evidenced by a greater reliance on measuring activity and on strategic thinking and, in a few cases, agency reorganizations to reflect the shape of programs). This tendency was more marked in some ministries and some programs than others. On the other hand, there was little progress in organically linking planning to budgeting. There were no evident criteria for the strategic programs selected, these programs were not effectively prioritized, and program

³ It is also our impression (not based on concrete analysis) that Axes of Regional Integration, the notion that underlay the indicative planning aspect of the PPA, did not really guide government actions in practice.

execution did not always follow the prioritization in the PPA, in spite of the cash-flow control.⁴

13. While the intention behind the PPA 2000-03 was in the right direction, we judged in our 2002 report that additional work would be needed to take fuller advantage of the PPA. In our judgment, such additional measures included: (i) introducing certain improvements in the technical design of the PPA; (ii) implementing additional public management reforms to support its consolidation; and (iii) somehow tackling broader external constraints especially the limitations to rational budget management imposed by the way in which Congress approaches budgeting and by fiscal adjustments.

- Improving technical design of the PPA included:
 - Sharpening of priorities;
 - Introduction of full program costing;
 - Introduction of in-depth program evaluations;
 - Monitoring of actual roles program managers play
- Additional public management reforms included:
 - Aligning organizational structures and incentives with the PPA's Program management logic;
 - Fostering greater ministry ownership of the PPA by, for example, involving sector ministries much more closely in the PPA program development;
 - Linking the PPA to more robust sectoral strategies, as a basis for designing PPA strategic programs and also for ministries' own performance management framework.

14. Already during the implementation of the PPA 2000-03, the government adopted some of these and other improvements. For example, the afore-mentioned cash-flow control regime can be seen as a means of sharpening priorities. Additional efforts were made to address some of the remaining issues in the PPA 2004-07.

The PPA 2004-2007: *Brasil de Todos*

Changes in the Model

15. For the large part, the new PPA maintained the model of its predecessors, but there were some changes.

⁴ The 2002 Bank report did not focus on the indicative planning aspect of the PPA (e.g., implementation of the guidelines developed in the *Eixos* study), but the corridor programs, though they have survived, have never really been prioritized. Since 2000, the sum of allocations for these programs in both the PLOAs and LOAs has typically fallen far short of the PPA allocations.

- *Planning.* In preparing the PPA 2004-07, the government undertook no new planning exercise such as the *Eixos* studying. As a result, the PPA 2007-07 may have lost some of its indicative-planning character that its predecessor was intended to have. Instead, the incoming Lula Administration formulated the Plan within a large exercise of public consultations at the state level, presumably to ensure effective coordination of federal programs with the states' priorities. Unfortunately, the absence of systematic follow-up seems to have discredited this exercise in participative planning.
- *Budgeting: prioritization of programs.* The modest steps taken to prioritize certain Programs through the cash flow control were reversed. There is said to be a degree of consistency between the PPA priorities and the *Metas Presidenciais* (Casa Civil), but many ministries have apparently not internalized these *Metas*, and the so-called inversion of priorities continues. The IMF-sponsored Pilot Program for Public Investments (PPI) has also appeared, more recently, as a "real" prioritization device of the government's public expenditure program, but its emphasis on projects may be somewhat contradictory to PPA's program logic.
- *Budgeting: medium-term perspectives.* In theory, PPA can strengthen budgeting by infusing a medium-term perspective in resource allocation decisions. Aware of the limitation that arose from the "static" nature of the PPA (i.e., a fixed rather than a rolling four-year plan), SPI began to include as an annex to the annual evaluation report a rolling plan (three years of expenditure are projected forward from the current year) as an indicative management tool (as opposed to the plan that is legally sanctioned by Congress). But these projections are not necessarily consistent with the forward projections of the LDO, nor are there robust forward estimates at the program level based on realistic estimates of projected program costs. In short, it is far from clear whether these indicative rolling plans actually guide budgetary decisions in any way.⁵
- *Budgeting: performance budgeting.* With the array of performance-related information (e.g., program objectives, performance indicators, annual evaluations), PPA has potential to promote performance orientation in annual budget decision-making. With the intent of encouraging Congress to look at the evaluation report together with PLOA (submitted on August 31), and thus to consider program performance as a criterion for resource allocation, the deadline for submitting the annual evaluation report to Congress has been delayed from April 15 to September 15. However, growth in current expenditures and the requirement for a higher primary surplus have led to tighter adjustment on discretionary spending, including a bulk of PPA priority programs. The share of the discretionary budget has continued to shrink. Non-earmarked revenues fell from 14 percent in 1996 to 7 percent in 2006, according to data of the Ministry of Planning. As an additional measure, SPI is considering making the production of an evaluation report a necessary prior condition for an agency's budget submission to be discussed by SOF/MP.

⁵ This is our conjecture as we were not able to extend out assessment to reviewing the role of the indicative rolling plans in budgetary decision-making. However, in none of the meetings we held with SOF and SPOAs, was the role of the indicative rolling plans emphasized by our interviewees.

- The *management model*. The new PPA includes several potentially important changes (Decree 5233, October 6, 2004 – summarized in Table 1). First, each program was required to prepare an annual Management Plan. Second, several steps were designed to provide greater alignment between programs and organizations. The Program Manager was required to be a senior appointment with authority to oversee the ministry unit (e.g., secretariat) in charge of a given program, aided by a more junior Program Manager and by Coordinators for each Action. The Program Managers and other senior officials were to come together in a Program Coordination Committee for each sector. A Program Management Committee would unite Managers and Coordinators of inter-sectoral programs. Third, a PPA Evaluation System was created in order to institutionalize M&E.⁶

Table 1: Organizational Changes under Decree 5233

Function	Task
<i>Plano gerencial</i> (each program)	The management plan (including evaluation plan) for a program.
<i>Gerente de Programa</i> (each program)	Responsible for program management and naming <i>Gerente Executivo</i>
<i>Gerente Executivo</i> (each program)	Supports <i>Gerente de Programa</i>
<i>Coordernador de Ação</i> (each action)	Manages Action
<i>Comitê de Coordenação dos Programas - CCP</i> (each sector)	Coordinates management processes to reach sectoral objectives by validating management plans for each program
<i>Comitê Gestor de Programa - CGP</i> (each multisectoral program)	Monitors and evaluates intersectoral programs according to their management plans.
<i>Sistema de Avaliação do Plano Plurianual</i>	System (CMA, UMAs) coordinated by MP
<i>Comissão de Monitoramento e Avaliação - CMA</i>	Propose M&E rules, processes; provide TA to results-based \$ resource allocation and program revision.
<i>Câmara Técnica de Monitoramento e Avaliação – CTMA *</i>	Provides technical support to the Comissão de Monitoramento e Avaliação.
<i>Unidade de Monitoramento e Avaliação - UMA</i> (each sector)	Supports the preparation of management plans and program M&E and provides TA on concepts and processes.

* Not part of Decree 5233

16. These changes were motivated by the SPI's realization of the need to improve alignment between the PPA program structure and each ministry's organizational structure as well as to strengthen the institutional bases of the annual evaluation exercises. These changes should have improved the PPA's effectiveness as a performance management tool. In fact, the Bank's previous report made similar suggestions as possible ways to improve the PPA. As we will discuss in greater details below, these measures have yet to translate into tangible improvements in the PPA's effectiveness. It seems that part of the reason has to do with the decline in government support for the PPA, despite the SPI's continued efforts to fine-tune the model.

⁶ SPI has expressed the intention to start selective evaluations within a central policy framework. Rapid evaluations would take up to six months to complete, at a cost not exceeding \$100,000.

3. THE EFFECTIVENESS OF THE PPA

17. The SPI attempted to fine-tune the PPA model by strengthening its management model and evaluation framework via Decree 5233. It might be argued that it is still too soon to conduct a thorough evaluation of the effects of these measures. However, after more than half a decade since its inception, the PPA has clearly passed its “honey-moon” period during which certain lack of understanding and resistance to internalization would be expected. Therefore, we interviewed a spectrum of sector ministry officials to gather their perspectives on the PPA as it has evolved over the years and as it currently functions. Our principal interest in the interviews we have conducted was to better understand the ways in which the PPA has improved government performance (with particular emphasis on the role of the annual evaluations). We expected the main channel for improved performance to be through the new management model which, it was hoped, would work through two channels: promote a focus on results within the agencies *and* allow the SPI to monitor the performance of agencies and mitigate barriers to performance. This section presents our general assessment of the effectiveness of the PPA as a whole in making the federal bureaucracy more performance oriented. A review of the PPA evaluation framework will be presented in a separate section below.

Box 1: Collecting the Evidence and Interpreting Perceptions

Methodology

An exploratory mission in February 2006 met with a broad range of government agencies with a role in evaluation. We had several meetings with SPOA officials and line managers responsible for programs in two ministries, Education (MEC) and Work & Employment (MTE). We also discussed evaluation and PPA issues with three groups of officials responsible for working with or in different ministries: SPI monitors, SOF analysts, and SPOA personnel. We also had meetings with TCU, SPI, IPEA, SEGES, Casa Civil (Subchefia de Articulação e Monitoramento – SAM), and CGU.

A mission in May concentrated on meeting with five ministries: Agriculture, Tourism, Transport, National Integration, and Justice. The choice of ministries was based on our discussions during the February mission. Our most important criterion for choosing these ministries was to represent a mix of those ministries reputed to be successfully following the PPA model and those that had less success or resisted it. (Given the often negative views on the PPA that we encountered in our February mission, we were particularly concerned in May to track down ministries with “success stories”, and our sample was probably biased towards these.) This is also a mix of large and small and new and old ministries. We met with SPOA officials and line managers responsible for programs and departments and with some Executive Secretaries. We conducted open-ended interviews to collect ministry perceptions about the PPA and evaluation. The main themes we explored were:

- PPA programs: their coherence, concordance with organizational structure of the ministry, and usefulness to ministries;
- PPA evaluations: process, quality, and the uses they were put to;
- The impact of PPA; and

- The management practices of the ministry: planning methods, use of information, etc.

Interpretation

In interpreting what we heard in the ministries, we should be aware of sources of bias. One arises from the role of ministries vis-à-vis the center of government, another from the particularities of different Public Service careers. Here are some of the general tendencies we detected from the interviews.

- We noted a tendency among sector ministry staff to see the PPA as an extension of the budget process rather than an instrument for performance management. Since the budget is allocated on the basis of PPA programs, it is little wonder that the PPA relationship between ministry and SPI is dominated by budget battles, rather than performance issues.⁷ As one of them put it, “The ministry plan is strategic; the PPA is budgetary.”
- Ministries tend to externalize their problems and expect external solutions to internal problems. For instance, budget quotas and delays in release of funds externally impose on ministries a critical obstacle to budget execution, but as better-managed ministries like Tourism have discovered, internal management can improve budget execution (even if it cannot eliminate the problem). In this sense, the self-evaluation is likely to highlight external constraints to program management rather than feasible improvements in internal management.
- SPOA personnel in the ministries tend to show more enthusiasm than their ministry colleagues for the PPA. This may reflect a career or professional bias. They, rather than line personnel, carry the burden of administering the PPA. They are part of a career that administers budgets and plans. Their job is to liaise between sector ministry and the Ministry of Planning, but their prime loyalties are to their Career (which will move them between ministries and between sectors and the center), rather than to the ministry. Line personnel in the ministries tended to view SPOA personnel as agents of intermediation with the Ministry of Planning, rather than of performance improvement.
- The interviewees often failed to provide specifics of how the PPA contributes to improved performance. Almost everyone interviewed, even those critical of the PPA, would still state the importance of PPA. But follow-up questions failed to evince specific examples of current benefits, indeed anything more specific than replies to the effect that “it’s important for sectors to know the government’s priorities” or “it’s important to promote better performance.” In some cases, we tried to “test” the meaning of such replies by asking: “if the PPA ceased to exist today, how would this affect the ministry’s life?” Responses ranged from “no effect” to a generalized reiteration that the “PPA was important.”

Some of our interlocutors were no doubt paying lip-service to the PPA because this is official policy. Others were probably expressing a more genuine belief in the importance of performance and perhaps reflecting that the PPA had initially helped promote this.

⁷ The only reported use of the annual evaluations was in the budget cycle. So what other purpose might ministries see in these evaluations other than to defend a budget position?

Integrating PPA Programs and Agency Structures

18. At the heart of the PPA management model is a system of input-output linkages and of accountability that, to function properly, requires an approximation of the agency's organizational structure and PPA programs (either through a direct correspondence in the organigram or through some form of matrix organization). In the program format, a single "problem" and a single product to solve it are identified, the product is associated with a number of specific actions to produce it (a form of "program logic"), and an information system helps in implementation and monitoring. Somehow reconciling these potentially conflictive logics of work organizations was one of the apparent intents behind the Decree 5233.

19. Simply put, aligning the program logic with the organizational logic could take the form of either the organizations re-structuring themselves to fit the program-based logic or the programs being redefined to fit the existing organizational structure. When we carried out our initial assessment of the PPA model in 2000-01, we came across some ministries that appeared to be embracing its philosophy, and even re-organizing the ministry structure to better fit the PPA's program logic. Thus, the Ministry of Transport had divided itself between those units that managed specific inputs in a traditional fashion (e.g., roads, waterways) and those that integrated them into multi-modal ("Corridor") programs (Secretariat of Development). Similarly, the Ministry of Environment had just re-organized itself into secretariats in charge of different environmental agenda ("green", "brown", "blue"), and explained to us how this new way of organizing itself was consistent with the PPA's program logic. The Ministry of Justice also made substantial efforts to align PPA programs and ministry structure, but has found that the programs were not well enough defined, in terms of identifying real problems, to be useful (Garcia, 2004). In all these cases, we met with enthusiastic PPA supporters within the ministries who seemed to see in the PPA a welcome impetus for performance-oriented changes to the ways the ministries organized and managed their own business. On the other hand, we also perceived resistance to change in ministries such as the Ministry of Health. We concluded at that time that this was probably expected since more institutionalized ministries such as Health were naturally averse to any kind of change.

20. It is not clear that the number of ministries actively adopting the PPA model has increased since the early days. Some ministries have been able, more or less, to approximate programs and organizational structures. For the ministries we interviewed, this was the case with Agriculture and Tourism. These ministries generally share the view that the program format has brought the benefits of clarifying objectives and providing a useful organizing principle. These ministries also seem to have moved towards more performance-oriented forms of management emphasizing planning and monitoring (see below).

21. At the other end of the spectrum are ministries that have resisted the adoption of programs in the spirit of the PPA. In the process of negotiating programs with SPI, these ministries have ended up with programs that are unsatisfactory both to them and to SPI. As a result, the PPA program plays no effective role for these ministries. They manage

on the basis of organizational structures unaffected by the PPA and they budget on the basis of Actions consistent with organizational structure. They prepare budget requests and report to SIGPlan simply by aggregating the data of Actions, but their activities are not necessarily guided by the pursuit of results as defined in the PPA program format.

22. A number of ministries are thought to fall into this latter category, judging by views we heard through our interviews.⁸ The Ministry of Education (MEC) is perhaps the most prominent example. According to the ministry officials we interviewed, there are three reasons for resisting the PPA concept of programs. First, programs that reflect the traditional organization of the ministry by different levels of education make more sense to MEC. (The alternative would be programs oriented to problems such as quality of education.) Second, the ministry prefers large programs because this gives it budget flexibility. Third, the ministry would like some programs which reflect issues that resound with politicians and the public – such as “school meals” – instead of programs that tackle central problems in education (such as student learning). Conceivably similar reasoning would easily apply to other ministries.

23. Of course, it may also be that ministries resisting the program approach have simply wanted to continue to do business in traditional ways, even if these are proving ineffective. This might be particularly the case for ministries more strongly subject to external political influences or to their own internal corporatist interests. Other ministries (National Integration, Transport) have found the program approach useful within individual Secretariats, though the ministry as a whole has been less enthusiastic. For example, the above-cited Corridor programs of the Ministry of Transport have remained in the PPA 2004-2007, but are no longer of relevance to the way the ministry does business in the sense that the commitment to multi-modal transport appears no stronger today than 5 or 6 years ago (or possibly weaker).

Internal Management

24. Some of the ministries have begun to adopt tools of results-based management, clarifying objectives through an analytical and planning process and developing information systems to monitor the process of implementation. Among the ministries we interviewed, Agriculture and Tourism appeared to have made progress in implementing planning and information systems. Education has made some advances, though so far less systematically. The Ministry of Labor & Employment has some elements.

⁸ Even in 2000-01, when we conducted our first review of the PPA in its initial moments, we heard complaints from program managers that actions that they did not consider natural or legitimate parts of their programs were forcefully inserted in them. At the time, we judged that this was both a reflection of natural resistance to change and a “growing pain” of not being able to get everything technically right from the beginning. The implication of such a judgment is that in an optimistic case this resistance and “growing pain” would lessen or disappear over time. If anything, our impression is that these problems have worsened, at least in some parts of the government. Although we have not done a full survey of the evolution of program structures in ministries, we have noted that in some cases what used to be programs have since become actions within a larger, umbrella-type “program (e.g., Family Health).” We see this as a retreat by the PPA.

25. Obviously, these tools are consistent with the management model of the PPA, but the extent to which they have developed *because of* the PPA is less clear. The preparation of the PPA 2000-2003 in 1999, then the PPA 2004-2007 in 2003, required agencies to plan at the sectoral level as an intermediate step between the PPA's Programs and its Macro-objectives. This initiated a serious planning process in some ministries (Justice, for example, though this ministry has not been able to monitor plan implementation effectively). In other cases, management improvements seem to have come about independently of the PPA. In the Ministry of Health, for instance, better information systems were required by the need to oversee the implementation of reforms in sub-national governments within the Unified Health System (SUS) framework. Similarly, the Ministry of Education (like Health) developed information systems in order to aggregate information from many decentralized production units (schools and universities). More recently, its new management information system was developed, in part, to rationalize the provision of information to external government agencies, including SIGPlan, and only thereafter started to be used for internal management.⁹

26. The new PPA management model introduced with the Decree 5233 was intended to strengthen ministries' internal management by (i) empowering the program managers and (ii) making program management more consistent with each ministry's organizational structure. The interviews revealed very little evidence of effective adherence to this new management model. Anecdotes included a case where a senior ministry official was oblivious to the fact that he himself was now the program manager, or a new unit head who simultaneously served as the manager of the single program that his unit was responsible for apparently went about his business without special concern for the PPA logic of results focus.

27. Re-aligning PPA programs to the ministries' organizational realities is a sound approach to enhancing effectiveness of the PPA. But in some cases at least, this may have brought the problem "back to square one." If ministries or specific secretariats are not themselves performance oriented, the ability of the PPA to make them so is limited, as it does not offer the ministries some of the most critical management tools such as greater flexibility in human resource management and greater agility in procurement.

External Accountability

28. Programs have perhaps been more unambiguously useful for the purpose of external oversight than for internal management of ministerial activities. The CGU and the TCU substantially use the programs to select target areas for audit, while the System of Presidential Goals (Casa Civil) selected its priorities using PPA programs as its basic unit of measure. This provides some *prima facie* indication that the programs tend to make sense as units. In fact, the TCU supports the idea of reducing "umbrella" programs – PPA programs typical of large ministries like Education and Health that contain a number of more genuine programs.¹⁰ One interviewee told us that "while the designation

⁹ Decree 5233 of 2004 mandated that each program prepare an annual management plan (Plano Gerencial), but this does not appear to have become an important routine.

¹⁰ Re-allocations across programs require re-authorization from Congress, whereas re-allocations across actions within the same program – up to a certain limit – do not

of PPA programs was not optimal, it has helped us do our work,” especially as there has been some improvement in the quality of data associated with the programs. It is in its use by other oversight agencies that the PPA has proven its contribution to transparency. Even then, there is little evidence that, so far, Congress or the public have benefited from greater clarity in the presentation of government activities. By the same token, media reports on budgetary issues, which are fairly abundant, usually do not use PPA programs as basis for analysis.

29. In contrast to the pro-active stance by the audit bodies, the SPI is seen as a rather passive administrator, rather than a problem-solver, in the PPA context. A dominant view among the ministries, even those having more strongly embraced the PPA, is that the SPI is substantially inflexible in its administration of the PPA. The sense is given that SPI first imposed the programs – if only by virtue of imposing a methodology – and has then been unwilling to change them.¹¹

30. Another virtually universal view was that SPI fails to act to alleviate the external constraints, such as delayed releases of authorized budgets and inadequate staffing, that the ministries communicate to them through the evaluations. In a few cases, officials acknowledged help they had received from individuals in SPI, but in general SPI is seen primarily as a requestor of bureaucratic information rather than ministries’ ally and problem-solver.

Indicators and SIGPlan

31. One of the obstacles to the PPA is the technical difficulty in managing performance indicators. Even in the cases of the ministries that had successfully adopted the program approach, problems of measurement (e.g., setting up useful indicators) were endemic. Some things are impossible, or very difficult, to measure. In some cases, it was difficult to collect existing data from distant outposts (e.g., the ministry’s own regional office) or from entities not under the ministry’s control (for instance, data from sub-national governments and even from agencies attached to the ministry). In other cases – typically but not only in investment projects – results could only be measured after a long gestation period.

32. There was virtually universal frustration at SIGPlan among interviewees from line ministries. It was considered of limited value and difficult to use (as one interviewee put it, “not tailored to our ministry’s realities”). This reflected two main concerns. First, SIGPlan was unable (nor was it intended) to accommodate ministries’ particularities – in particular ministries where Actions were executed outside the ministry or had long gestation periods. Second, SIGPlan was not detailed enough for ministries to use as their

¹¹ In our ministry interviews, we found one interesting exception to this view: an apparently progressive and clear-sighted manager who had found little difficulty in substantially renegotiating the content of a program to align it with the ministry’s actual practice (but not getting extra money).[0] While we can imagine other similar cases, we did not come across any other example of this nature during our interviews.

own management information system (InfraSIG). In short, SIGPlan is for the ministries an immediate cost with few or no internal benefits.¹²

33. Decree 5233 of 2004 set up various mechanisms to strengthen the position of PPA programs and monitoring in the sectors: program managers were to be appointed from among more senior ministry managers; senior ministry managers would constitute a Program Coordination Committee (CCP) to coordinate and improve program execution; and a Monitoring and Evaluation Unit (UMA) would support PPA processes in the sector, especially M&E. The new program managers have been appointed and the UMAs are beginning to function effectively, though the CCPs have made little progress. These changes appear to have had no effect on the propensity of ministries to embrace the PPA approach to performance management or to resist it. Where there has been an effect, it has mostly been to improve the ministry's ability to feed information to SIGPlan and SPI.

34. The idea of the UMA was inspired partly by the Canadian model where line departments have an evaluation unit to advise the departments' chief administrators (vice ministers). In the Canadian model, each department has structured its evaluation unit (often but not always combined with an internal audit unit) to suit its internal management needs. The main objective of these internal evaluation and audit units is to support the chief administrative officers of the departments in their day-to-day internal management functions. In contrast, the UMAs are being used, according to our interviews, only for external reporting, largely de-linked from whatever measure the ministry may be taking to strengthen its internal management. Where ministries have established M&E activities as tools of internal management, they have done this separate from PPA structures. This is the case with the Ministry of Education and the Ministry of Social Development (MDS).¹³ These are telling indications of how the ministries perceive the PPA not as an instrument for improving their management but rather as a framework for reporting (to the MP) and resource allocation (both of which are seen as "necessary evils" for them to continue operating).

Conclusions

35. The seven ministries we visited showed large differences in management capacity and in adherence to the PPA model. Yet there was a considerable consistency, across the spectrum of ministries, about how the PPA was now functioning. We think that the following inferences can be taken from these responses. These are likely to apply to government agencies more generally.

¹² Virtually the only case we heard of SIGPlan helping a ministry was that it provided a useful vehicle for collecting information from far-flung ministry outposts. But the information was to feed SIGPlan; it was not used in internal ministry management.

¹³ MDS is one of the few ministries that has created its own monitoring and evaluation unit, the Secretariat of Monitoring and Information Management (SAGI). SAGI is carrying out a large number of in-house evaluations, including those of high technical quality, and "strategic" monitoring of the ministry's key programs. The MDS created an UMA separately within the Sub-secretariat of Planning, Budget and Administration (SPOA) to deal exclusively with PPA monitoring and evaluations.

- *The PPA model originally helped promote a focus on performance in the ministries.* In particular, the PPA allowed greater transparency in the government's business. But the PPA was not the only path to good performance and its positive influence has been felt unevenly across ministries. Many ministries today still have not entered on that path. A secondary route of PPA contribution to better management has come through the systematic use of the program structure as a basis for audit work by the TCU and the CGU.
- *The PPA, as it operates today, no longer is a major contributor to promoting better internal management in the ministries, and instead has become a bureaucratic burden to them.* The ministries make little to no use of the PPA's monitoring and evaluation tools in their internal management. With a few exceptions, the efforts to align the PPA program logic and the organizational realities of the sector ministries have not resulted in performance-oriented ministries designing and managing programs with strong focus on results. Some of them do use evaluation results in budget negotiations, but the government (i.e., MP) has not established a clear policy to use evaluation results in the sense of performance budgeting. The ministries complain of getting little to no help from SPI to alleviate the problems identified in the evaluations and they regard the PPA as a largely bureaucratic procedure.

36. The main elements in the perceptions and judgments underlining these three points are summarized in Box 3.

Box 2: The Impact of the PPA: a Summary of Perceptions and Inferences

1. The PPA model originally helped promote a focus on performance.

- Some ministries have successfully aligned internal organizational structure and PPA programs and developed their own, complementary routines of planning and monitoring. This is leading to a greater focus on coherence and results.
- Other ministries have tried and failed to adopt the PPA model. Others have resisted it. Some bob-adopters are also developing their own routines of planning and monitoring, and they are also moving, perhaps less systematically, towards a greater focus on results.
- The program format, by making government business more transparent, has helped other government agencies, particularly the audit agencies, do their job.

2. The PPA, as it operates today, no longer promotes good performance, but is a bureaucratic burden.

- The ministries do not find SIGPlan a useful tool: it neither accommodates ministry-particular characteristics for monitoring nor provides enough ministry-particular detail for planning.
- The ministries do not use program evaluation results for internal management, though the ministries and SOF do use evaluation results as one extra tool in budget negotiations.
- Ministries perceive SPI as planning theorists who imposed programs and then proved inflexible in modifying them.
- Ministries say that SPI fails to act to alleviate constraints that are identified in the evaluations, but the audit agencies, through impositions or advice, have greater impact on ministry management.
- While ministry officials still say that the PPA is important, they are unable to provide concrete examples of its positive contribution.

4. EVALUATION UNDER THE PPA

37. Monitoring and evaluations (M&E) have been an integral element of the PPA model from its inception in the PPA 2000-03. But the emphasis has shifted from the near exclusive focus on “real-time” monitoring via SIGPlan in the PPA 2000-03 to the increasing emphasis on program evaluation in more recent years. In the original design, the annual evaluations were designed as “meta monitoring” whereby the program managers themselves report on consolidated program performance at the end of the year. This initial choice has had consequences for how the system has evolved and performed, and how it has come to be seen by line ministries.

38. The PPA annual evaluations take place within the broader evolving context of PPA implementation assessed so far. The broader context is critical for understanding ways in which the evaluation framework can be improved. First, the design of the PPA evaluations closely follows the overall design of the PPA model, and any adjustment to this model would have direct implications for the conceptual design of the evaluation model. Second, the effectiveness of the PPA evaluations hinges on the effectiveness of the overall PPA model. It is highly unlikely, at least so long as conceptual consistency between the overall model and the evaluation framework is maintained, that evaluations can be more effective than the overall model itself.

39. Given the initial objective of this assessment, we provide a separate assessment of the PPA annual evaluations in this section. However, our assessment is embedded in the overall assessment discussed so far. In the final section, we offer alternative ways of thinking about ways to strengthen both the PPA model in general and specific approaches to evaluations.

The Ex-ante Objectives of Evaluation

40. The declared objectives of PPA evaluation, as is often the case in other countries, are general and comprehensive. For instance, the 2006 Manual (page 7) lists: transparency and accountability; support to decision making in planning and executing government activities; helping organizations to learn about their programs; and improving the design and management of plan and programs. Congress gets no particular mention.

41. The intended objective can also be inferred from the questions asked in the evaluation routine. The topics of the 2006 Manual (which is evaluating performance in 2005) are listed in Table 2.

- Some 16 pages of questions examine results, design, and implementation at the *program* level. “Implementation” is largely concerned with identifying the adequacy of inputs and of internal and external constraints to performance.
- Some 16 pages are devoted to results, design, and management at the *sectoral* level. (“Profile” questions also enquire whether/how PPA-related institutions and systems are working.) Evaluation of sectoral results and design is somewhat summary, but

there are 13 pages of questions on sectoral management (most of this being a specific set of questions developed by SEGES/MP that were more detailed than the sectoral management questions typically asked in the annual evaluation routine).¹⁴

42. To judge, overall, from the questions posed, evaluation concentrates on program results (and the indicators measuring these), constraints to program management, and the characteristics of sectoral management. There is little evident emphasis on program impact (effectiveness and relevance), program efficiency (inputs per output), or inter-sectoral expenditure prioritization.¹⁵ In the evaluation of program design, the emphasis is more on identifying specific lacks and needs, rather than on understanding how the program is meant to work. The use of standardized questions impedes a greater emphasis on a program's logical framework.

The Evaluation Process

43. The PPA annual evaluation process is perhaps one of the most institutionalized among the countries that purport to have government-wide M&E systems, along with the Chilean system of "Management Control."¹⁶ Practically all finalistic programs in every sector of the federal administration (and the TCU) have been evaluated every year since 2000 (and most of these evaluations have been published). In the 2005 annual evaluation report, 363 out of a total of 394 programs were evaluated.¹⁷ Program Managers and SPOAs respond to a standard set of questions (see above). These questions are largely closed-ended or multiple-choice. Responses are largely based on perceptions or subjective judgment of the managers, but financial and physical execution data from SIGPlan complement some of the information (see the separate analysis reported below on the content and the quality of a sample of program evaluations).

44. Evaluation is linked in a sequence with budget preparation (Table 3). The first stage of the sequence is program and sector evaluation, which now occurs from February to May. First, the managers evaluate their programs. Then the sectors evaluate the programs and sectoral management. MPOG reviews the evaluations, discusses them with the sectors, and then finalizes them.

45. During the time of program and sector evaluation (February to March), MPOG (SPI, SOF, and IPEA) is able to monitor, via SIGPlan, the progress of manager and SPOA responses. SPI analysts can also intervene in the process, for instance, to tackle problems of inconsistencies in the responses. Beginning in March SPI reviews the resulting draft evaluations to prepare its own draft comments which it then discusses with SOF, IPEA, and the sector. This contributes to an MPOG budget document, *Ficha de Avaliação de Programas*, containing MPOG's view of the program. This is a document

¹⁴ The questions on sectoral management are largely based on the methodology of organizational assessment of the Programa Nacional de Gestão Pública e Desburocratização (Gespública).

¹⁵ A similar point was made in World Bank (2002, pp. 33-34).

¹⁶ The PART framework recently developed in the US is another centrally-driven, well-institutionalized evaluation process.

¹⁷ These included 34 *Programas de Gestão de Políticas Públicas* in each ministry, and excluded 13 *Programas de Serviços ao Estado* and 3 *Programas de Apoio Administrativo*.

with qualitative information which goes to Congress before MPOG has announced to the sectors their spending ceilings (i.e. prior to the quantitative stage of the budget).

Table 2: The Topics of PPA Evaluation, 2006

Avaliação do Programa (Etapa Gerente)	Avaliação do Conjunto dos Programas (Etapa SPOA)
<p><i>I Avaliação dos Resultados:</i></p> <ul style="list-style-type: none"> • Results achieved, compared to expectations. • Relevance of indicator/ absence of indicator. • Coverage of target population. • Beneficiary satisfaction. • Existence of other evaluations? 	<p><i>I Avaliação dos Resultados:</i></p> <ul style="list-style-type: none"> • Results achieved. • Principal success factors. • Main difficulties in getting results. • Contribution of main results to government strategy.
<p><i>II Avaliação da Concepção:</i></p> <ul style="list-style-type: none"> • Reasons for inadequacy of design • Areas for improvement. 	<p><i>II Avaliação da Concepção:</i></p> <ul style="list-style-type: none"> • Consistency of sectoral strategy with strategic orientation of government. • Consistency of 2005 budget programming, 2005 modifications, and proposed modifications with sectoral objectives. • Impact of sectoral policy on transversal issues (race, gender, disabled, youth).
<p><i>III Avaliação da Implementação:</i></p> <ul style="list-style-type: none"> • Characterize monitoring mechanism. • Physical goals achieved, compared to LDO. • Sufficiency of budget execution; compatibility of money flow with physical programming; impact of non-budget money; impact of congressional amendments. • Share of administrative in total costs; uses of administrative costs; implications for rest of program. • Adequacy of materials/infrastructure; needs. • Adequacy of human resources; needs. • [External] restrictions on performance. • Programs with decentralized execution: performance; information; manager's role in execution. • Multisectoral programs: performance, constraints. • Actions executed in other parts of ministry: performance; information; integration with rest of program. • Partnerships: performance of partners; constraints. • Social participation: mechanisms; contributions; constraints. • Beneficiary satisfaction: mechanisms; constraints. • Good management practices that are replicable. 	<p><i>III Roteiro de Avaliação da Gestão Setorial:</i></p> <ul style="list-style-type: none"> • <i>Profile:</i> functioning of CCP, UMA, SIGPlan; role of MP. • <i>Functioning and results of management model (SEGES):</i> <ol style="list-style-type: none"> 1. <i>Coordination:</i> relationship of sectoral organization & management to program management. 2. <i>Sectoral formulation of strategies:</i> <ol style="list-style-type: none"> 2.1 Strategy formulation. 2.2 Development of strategies into programs. 3. <i>Relation with beneficiaries and society:</i> 4. <i>Information management.</i> 5. <i>Operationalization of sectoral actions:</i> <ol style="list-style-type: none"> 5.1 Management. 5.2 Managing people. 5.3 Budget & financial management. 5.4 Management of supplies. 6. <i>Results:</i> <ol style="list-style-type: none"> 6.1 Support processes. 6.2 "Finalistic" processes.

Source: *Manual de Avaliação do Plano Plurianual 2004-2007, (Exercício 2006, Ano Base 2005)*

Table 3: Timetable of Evaluation and Budget Preparation

February	March	April	May	June	July	August	September
Program & Sector Evaluation				Budget preparation (quantitative)			
						Evaluation Report	

Source: Anexo D of Manual de Avaliacao

46. The quantitative phase of budget preparation follows immediately, from June to August. SOF/MP announces spending ceilings for each sector, the sectors then make detailed proposals which MPOG then analyzes. MPOG then delivers the draft annual budget law (PLOA) and the draft law for the annual revision of the PPA (i.e. rolling PPA) to Congress by August 31.

47. Finally, SPI/MP prepares the Annual Evaluation Report from August onwards and delivers it to Congress by September 15. In preparing the Annual Evaluation Report (August-September), analysts from SPI and IPEA provide some critical analysis of the managers' program evaluations. This provides the basis for a text on each Program which is discussed with the sectors. In the process of negotiating a text, our interviews suggest that the analysts do not usually change or add much.¹⁸ Thus, this is essentially a self evaluation by program managers.

48. Over time, there have been changes in topics covered (notably, the theme of management getting more or less attention), in timetable (presentation of the Report to Congress was brought back from April 15 to September 15), and in conceptual orientation (2002). But in general, the overall design of the evaluation exercise has stayed basically the same, while Annual Reports seem to have differed mostly in the amount of time that was spent preparing them.

The Quality of Evaluation

49. To assess the quality of the PPA annual evaluations, we reviewed a small sample of programs for the 2000-2004 period (for a program-by-program analysis, see appendix 3).¹⁹ The sample covered nine programs from some of the ministries we visited during our two missions. One of the criteria for selecting programs to be reviewed was whether the program had been evaluated by the TCU as part of its operational audits. The comparison with the TCU evaluations did not necessarily imply an assumption that these were of better quality than the PPA annual evaluations. But rather, we judged that the TCU evaluations, which are conducted based on a more or less standardized

¹⁸ Since the law only requires reporting of physical and financial execution, SPI has to obtain the voluntary cooperation of managers and SPOAs in the evaluation process. There were some suggestions in the interviews that SPI's relationship with program managers (especially of strategic programs) has become less close – this is perhaps due to fewer SPI resources – while SOF's relationship remains closer.

¹⁹ When the review was conducted, the 2005 evaluations were not yet available.

methodology, served as reliable comparators to identify any systematic advantages or disadvantages inherent in the methodology used for the PPA annual evaluations (and vice versa).

50. Each program evaluation was assessed subjectively according to seven criteria:²⁰

- whether the evaluation report presented a clear description of the program and its performance;
- how the recommendations changed or were repeated over time;
- whether concrete solutions to the identified problems were suggested;
- whether the evaluation report contained verified performance indicators as originally defined in the PPA;
- whether the evaluation made use of other indicators;
- whether the evaluation drew from other external evaluations; and
- whether the findings are consistent with the findings of the TCU evaluation.

Clarity of Evaluation Reports: Our assessment shows that the descriptive clarity of the annual evaluation reports is satisfactory, but the quality varied across programs in most of the other dimensions. Firm generalizations are not possible given that the sample is not representative of the entire set of annual evaluations. In general, the self-evaluation reports enable the reader to understand, based on the program manager's opinion, how the particular program works and what difficulties it has faced. But in our small sample, only two programs (*PETI, Saúde da Família*) had consistently good quality along the seven dimensions.

Evolution of Recommendations: Only in a couple of the programs (*PETI, Saúde da Família*) was there a sign that problems raised in a previous year had been resolved by the following year. In some cases (*Novo Mundo Rural, Turismo no Nordeste, Deficiência*), the same problems were repeatedly raised in the successive annual evaluation reports, while in others corrective measures were taken only in response to the TCU evaluation or by change of government (*Energia nas Pequenas Comunidades, Educação de Jovens e Adultos, Morar Melhor*), suggesting that the PPA annual evaluations served no purpose in inducing such corrective measures.

Concrete Solutions to Problems: Five programs lacked concrete solutions to the problems identified in the annual evaluation reports. Usually managers complain of certain problems, but are not able to come up with solutions for their problems. There are also recurrent issues that are never addressed, such as: discontinuity of financial flow due

²⁰ A more systematic approach, for example, one based on clear criteria for rating the evaluation in each of the seven or other categories, would be desirable. For our rapid assessment, we were not able to develop such a methodology.

to *contingenciamento*, problems of coordination with other implementing ministries, lack of infrastructure and personnel capacity.

Measure of PPA Indicator: Only one program (*Saúde da Família*) was able to measure the indicators in the annex of PPA report. All the others were incomplete. Most managers complain in the text that the PPA indicator was inadequate and/or impossible to measure. During interviews, several managers mentioned the need to improve the process of indicator selection.

Use of other indicators: Several programs reported indicators other than those specified in the PPA, which is positive. Most indicators come from the national research institutes, like IBGE (Census and PNAD). In three programs of the sample (*Energia nas Pequenas Comunidades, Morar Melhor and Novo Mundo Rural*) there were no indicators at all. This is worrisome, because it is a sign that managers might be operating their programs “in the dark.”

External Evaluations: Some managers commissioned external evaluations of their programs (*PETI, Energia nas Pequenas Comunidades, Saneamento Básico, Turismo no Nordeste*), which is positive, but only two of them used the results in the PPA report (*PETI and Turismo no Nordeste*). These cases show, however, that the Ministry of Planning might be able to stimulate the use of external evaluations to help managers identifying problems and finding solutions.

Consistency with TCU Evaluations: Four of the programs “missed” serious issues identified in the TCU evaluations. For instance, for the program *Educação de Jovens e Adultos*, TCU captured high repetition rates in the literacy course, but this was completely omitted in the PPA evaluations. TCU recommended several changes in the program, such as increase in the length of the course and maintenance of better student records. The program was revamped both to comply with TCU but also because the new government took office. Some of the programs reports were coincident with TCU’s (*PETI, Saúde da Família, Morar Melhor, Saneamento Básico*), but even when the same issues were identified, the PPA evaluations inevitably approached them much more superficially because of its methodology.

The analysis of the quality indicates that information generated from these evaluations is unlikely to be of great value to the program managers. At limited cost, the quality could be improved, for example, by (i) using a broader set of methodological instruments to collect data from municipalities (mail questionnaires, interviews with specialists, focus groups with managers and beneficiaries, direct observation, etc.); (ii) commissioning external evaluations and use existing evaluations; and (iii) providing training in research methods for program managers. But to what extent and how the quality of these evaluation reports should be improved should be a decision based on the intended use of the reports. In the current form and content, we believe that the most obvious use of the PPA annual evaluations is public information and accountability about government work. This is itself an important purpose, but if the ultimate objective is to use evaluations to improve government performance, then the whole approach may require re-thinking.

Table 4: Summary of Evaluation Quality

Programa	Clareza da Avaliação	Evolução das recomendações	Soluções para problemas	Indicador PPA apurado	Outros indicadores utilizados	Avaliação externa	Consistente com TCU
PETI	✓	✓	✓	✗	✓	✓	✓
Saúde na Família	✓	✓	✓	✓	✓	✗	✓
Energia Pequenas Comunidades	✓	✗	✓	✗	✗	↔	✗
Educação de Jovens e Adultos	✓	✗	✗	✗	✓	✗	✗
Novo Mundo Rural	✓	✗	✗	✗	✗	✗	✗
Turismo no Nordeste	✓	✗	✗	✗	✓	✓	✗
Morar Melhor	✓	✗	✓	✗	✗	✗	✓
Saneamento Básico	☺	☺	☹	☹	☺	☹	☺
Atendimento ao Portador de Deficiência	☺	☹	☹	☹	☺	☺	☹
Resultados	✓	✗	✗	✗	✓	✗	↔

Note: ☺denotes positive assessment, ☹denotes negative assessment, and ☹☺ denotes mixed assessment.

The Uses of Evaluation

51. The ultimate value of evaluations is in their actual use in decision-making. As reported in greater detail below, our interviews offered very little evidence of effective or systematic use of evaluation results.

52. The potential usefulness of the evaluation process can be gauged, to an extent, from the coverage and quality of the Annual Evaluation Report. Except for the latest report (2006) which covers 2005 and includes a section on the “evaluation of the management of the plan” (Box 3), the emphasis is clearly on individual programs: there is little attempt to aggregate or generalize on findings (related, for instance, to constraints on performance or budget/plan prioritization) at the level of either the sector or the government. To the extent program performance suffers from common systemic constraints such as those that result from the nature of the government-wide policies over human resource management, financial management, and procurement, analyzing them in the aggregate and exploring remedies for the government as a whole (a natural role of the Ministry of Planning) would have been a valuable contribution of a comprehensive evaluation such as the PPA annual evaluation. Similarly at the sectoral level, ministry authorities could use the PPA program evaluations to identify specific ministry-wide issues that hamper management of the ministry programs. The focus on programs as nearly exclusive units of evaluation therefore limits the evaluations’ potential utility.

Box 3: The Evaluation of the Plan Management

The 2006 PPA evaluation report (about fiscal year 2005) includes an assessment of how the PPA as a whole has been managed. The analysis was done with the information collected from 369 programs about issues like human resources, logistics, information technology and financial resources.

PPA Management Model: From the universe of 30 agencies that responded to the specific questions, 25 affirmed use of the SIGPLAN to manage and monitor their programs. Eleven agencies (including Agriculture, Science and Technology, Social Development, Education, Planning, Health, Tourism) claimed to have their own INFRASIG. Twenty agencies regard MPOG’s assistance to programs as adequate.

Budget and Financial Flow: About 43% of managers pointed out that budget execution was sufficient to implement the program, mostly from Ministries that benefited from earmarking (Health, Education, Finance, Planning, Energy), whilst for 52% budget was insufficient. For Ministries like Agriculture, Transport, Environment, and Defense, the insufficiency rates varied between 65% to 94%. Likewise, the financial flow was considered compatible with programming for those ministries with earmarked funds (28% of the ministries). For 37%, there was discontinuity in fund releases without problems for program execution, but for 33% the discontinuity disrupted program implementation.

Material and Infrastructure Resources: Around 59% of Ministries revealed inadequacy in this topic. Problems related to lack of materials was concentrated in the Ministries of Defense, Transport, Justice and Health. Conversely, infrastructure problems affected all Ministries, but mainly Cities, Justice and Social Development and at the decentralized level in Agriculture,

Defense and Health. The most common issues are lack of IT equipments, laboratories to develop new technology, cars, installations, furniture, telephones and physical space.

Human Resources: Inadequacy of HR was considered to be a systemic issue for 75% of the programs due to reasons such as lack of a career plan, and high turnover of staff due to *concurso*s with higher salaries. Turnover is also responsible for deficient staff qualification, requiring continuous retraining. As a consequence, staff is composed by trainees, political appointees, consultants, and temporaries with inadequate qualification. Therefore, problems related to human resources are considered the most serious impediment to program implementation.

Restrições: The most cited shortcomings are: budget cuts (57%), delays in fund releases (44%) and difficulties in contracting and procurement (34%). Other difficulties are: the cumbersome process of acquiring environmental licenses, absence of legal rules. Most managers declared high execution of decentralized programs (45%), integration between executors and central government (87%) and availability of information from municipalities (81%). The performance of multisectoral programs has also improved with 65% of programs with high or medium levels of execution.

Social Participation and Beneficiaries Satisfaction: About 80% of programs have some form of social participation. The most common methods are: meetings with interest groups (50%), sector councils (29%), ombudsmen (29%), public hearings (21%), and public consultation (18%). Other methods are: Internet website, forums, management committee, and call centers. Regarding beneficiaries satisfaction, 59% of the programs do not have formal mechanisms to measure beneficiaries' satisfaction and 79% responded that the program did not have other evaluations, apart from the annual PPA evaluations.

53. MPOG sends the Report to Congress and publishes it on the Internet. As far as we know, it does not otherwise distribute it or publicize it (e.g., a press conference). Neither are the Report or the recommendations it makes for each program presented or discussed, with the ministries. Currently, MPOG compiles some information on program implementation problems (notably “*Restrições*” or external constraints) or on sectoral management problems. But it does not act on this information to propose corrective measures. MPOG’s failure to act on the information that the sectors have provided it on external constraints has led to some dissatisfaction on the part of the latter.

54. Program evaluations are supposed to feed into the annual program review process that takes place between the SPI and the ministries on a bilateral basis. As discussed below, however, this phase tends to be seen by line ministries as mere precursors to the subsequent budget negotiation phase vis-à-vis SOF, rather than an opportunity to discuss program performance and policy priorities.

55. According to some interviewees, there are instances in which evaluations are used in budget negotiations. This has happened both within sectors (i.e. in the preparation of a budget request) and in the sector’s negotiation with MPOG. But it is not clear if this occasional use of the annual evaluation results constitutes a variant of performance budgeting, whereby budget allocation decisions are influenced, at least on the margin, by information regarding program performance. The use of evaluation findings in the

budget formulation process appears to be rather ad hoc. In one internal MPOG/IPEA seminar on evaluation methodologies, our suggestion to consider using the PPA annual evaluations as a tool for improving budget allocation seemed to elicit little support among the ministry's and IPEA's technical-level staff. From these observations, we conclude that the use of evaluation findings in the budget process is limited and perfunctory. As one executive program manager put it, "The only relevance of evaluation is: if you don't work with PPA, you don't get budget."²¹

56. Overall, it was a virtually universal perception that no ministry – whether it embraced or resisted the PPA – professes to benefit from the evaluation process or makes use of the results of the annual evaluations for the purpose of improving internal management (efficiency, effectiveness, etc). PPA evaluations are often carried out by the SPOA, with few inputs from the rest of the ministry. In only a minority of ministries is there even enough interest to make a methodological critique of evaluation. This suggests, among other things, that evaluation is culturally foreign to most ministries. As with SIGPlan, ministries incur a cost, but see hardly any benefit.

The Outcomes

57. Overall, the system tends to produce weak, self-serving evaluations. Though some of the evaluations may be potentially useful, the system is, we believe, hardly likely to be cost-beneficial given that collection and transmission of monitoring data (to feed the SIGPlan) and annual evaluations impose transaction costs to ministries that are not negligible, especially in view of other centrally-mandated reporting requirements (e.g., from Casa Civil, CGU, TCU). The weak coordination among the central agencies referred to above exacerbates the overload of the reporting requirements for line ministries, and is generating some resentment among them. Of course, the point is not that line ministries should not have to report on their performance to central agencies such as the Ministry of Planning, but its transaction cost should be kept to a minimum by avoiding unnecessary duplication (a frequent complaint among line ministries) and its (perceived and real) benefits should be increased by effective utilization of the information provided in meaningful and visible ways (so that the line ministries would easily understand the rationale for providing information). Although understandably line ministries would rather have no reporting requirement to the center, that would not be a reasonable expectation. But if the information submitted is not seen to be leading to meaningful ends, the ministries would rightly question the whole purpose of the particular reporting requirement.

58. This indeed seems to characterize the ministries' perception of the information requirements associated with the PPA (i.e., updating of the SIGPlan and the annual evaluations). As some staff of the CTMA also admitted, the annual evaluations (partly because of its nature as self-evaluations) have become a "complaint box." Program managers use it to register their "complaints" and are repeatedly disappointed because the Ministry of Planning usually offers no response (let alone solution to their problems).

²¹ SPI maintains that the Ministry of Agriculture is using evaluation purposively, but gave no details (other than about the ministry's efforts to improve the information coming from its State-level offices).

Similarly, line ministries tend to perceive SIGPlan as only serving SPI's ill-defined monitoring function without providing the ministries themselves with concrete benefits – the data entered in the SIGPlan tend to be too aggregated to be useful for day-to-day program management and thus not very useful for program managers, while ministry authorities seem to make little use of this information.

59. One positive development that seems to have been induced by the SIGPlan was the movement to develop a ministry-level management information system (Infra-SIG) that can be linked to the SIGPlan. With some brokering by SPI, ministries are now exchanging information and experience with Infra-SIGs.

60. Nevertheless, it seems to us that such an outcome still leaves open a more fundamental question. What is the basic rationale for a central agency such as SPI to collect relatively aggregated information on program performance during the budget year? The initial vision was that the program managers would be in constant contact with SPI so that the latter would monitor program progress not only for reporting purposes but more importantly for problem-solving purposes. This latter objective seems to be neglected, and the whole SIGPlan apparatus has lost its most important justification. In the absence of active problem-solving by SPI, does it make sense for the Ministry of Planning to collect program performance data on an ongoing basis (which in reality means as often as program managers are willing and able to input the data, which varies from program to program)? Isn't reporting on a less frequent basis (e.g., annual) sufficient?

Conclusions

61. The system is characterized both by a potential problem of the evaluator's motivation (because the sector evaluates at the behest of MPOG) and by its design as self-evaluation (i.e., managers evaluate their own programs). This combination leads to several problems:

- Program managers have no motivation to be critical of their programs (except inasmuch as criticism might secure more resources);
- Program managers tend to lack the capacity to evaluate and would have little incentive to invest in building this capacity – everywhere program units prefer allocating scarce budget for program execution than evaluation, but the sense that this is an evaluation that is mandated (imposed) from above would exacerbate such an “anti-evaluation” perspective;
- Some sectors are not performance-oriented, thus lack the motive to evaluate and have no systematic approach to utilizing evaluation results when these are conducted;
- Others lack the information necessary for evaluation (c.f. the indicators problem); others politically resist the influence of MPOG. When evaluations are done selectively, availability of data is usually a key criterion for program selection. For those programs that lack adequate data, ministries or program managers are often instructed to develop sufficient data to permit evaluations in the future. The blanket-

approach of the current PPA annual evaluations lack differentiated treatments of programs depending on their “evaluability” status, although this is precisely one of the aspects the CMA/CTMA are trying to improve upon.

- Self-evaluations are less likely to offer program managers much informational value-added (i.e., they learn little from evaluations that they themselves conduct without in-depth additional data collection and analysis).

62. These factors, sometimes in combination, lead to low-quality and/or self-serving (defensive) evaluations.

63. Given the current design, we believe it is not surprising that ministries do not turn to evaluation information for the purpose of improving their own management, program designs or resource allocations. Although some of the information generated (especially the results of the “sectoral evaluations” as well as aggregation of program-level findings) could be useful at the high level within the ministries (e.g., Minister and Executive Secretary), from the interviews we heard of no case where the senior ministry authorities systematically utilize this information for their decision-making. This leaves SPI as the one actor with both an interest and an ability (at least in theory) to make use of the evaluation findings. But a common complaint was that SPI rarely acts on the evaluation findings to remove or alleviate constraints to program implementation (SPI may use the evaluation findings to engage in discussions on program revisions, but this itself seems to generate another set of problems as discussed below). The end result is that there is misalignment between the content of the information generated (tending to focus on external restrictions and provide “no news” for the program managers themselves), the “owner” of the process (those with the greatest interest in making the system useful, which in this case is clearly SPI), and what the “owner” is able/willing to do with that information (SPI is currently not able/empowered to remove these external restrictions to aid program implementation at the sector level). When these elements are poorly aligned, evaluations are unlikely to lead to effective utilization.

64. Finally, the self-serving nature of the PPA evaluations may be reinforced by linking evaluation closely to budget preparation, perhaps ironically. While linking evaluation to budgeting is both standard and perfectly reasonable, self-evaluations with limited external check on their robustness (despite the validation by SPI analysts) invite opportunistic use of evaluations for the primary purpose of making a case for more resources. An independent evaluation linked to the budget process could avoid such opportunism (e.g., Chile), although that would open up for the other side of opportunism (i.e., the budget authority, SOF in Brazil, picking up only negative evaluation findings to make a case for lower allocation, as some ministries seem to perceive is the practice already).

65. Even with more SPI resources and the current efforts to systematize evaluation (UMAs etc.), these design problems would still be likely to undermine the PPA evaluation system.

66. Overall, there is little evidence that evaluation is producing results that are used, hence contributing to improving program (thus government) performance. The failure of SPI and the disinterest by line ministry authorities to use information provided in the evaluation process have reduced the credibility of evaluation. As discussed previously, however, this is just one aspect of a set of problems that undermine the credibility and thus the effectiveness not only of the PPA evaluations but also the PPA model as a whole within the federal administration.

67. How can the annual evaluation process be improved? Our assessment of the evaluation quality suggests that there are likely to be low-cost ways of improving the reporting format as well as the informational content of the annual evaluation reports without changing the current basic design. Even if the self-evaluation approach is maintained, willing program managers could actively seek supplementary sources of data and analysis, such as existing external evaluations or by commissioning a simple beneficiary survey, for example. But a real question is whether the program managers themselves see a reason to expend a little extra effort to improve the quality of their annual evaluation reports. Based on our observations so far, such an incentive does not seem to exist at least among average program managers.

68. As in the case of the PPA model as a whole, the poor outcome of the PPA evaluation process can be related to some design weaknesses in the evaluation system. Like the PPA more generally, the evaluation system is too universal and too standardized. Brazil is the only country in the world to put all of government actions into programs and evaluate every one of these programs every year. SPI has long recognized this particular weakness, and has been in the process of developing a framework (policy and methodological guidelines) for selective, more in-depth program evaluations (without necessarily abandoning the universality at the same time). This is a welcome development and is expected to mitigate some of the flaws in the PPA evaluations discussed above. But unless improved evaluation methodologies are accompanied by better incentives and capacities for utilizing and acting upon the findings, the credibility of the revamped evaluation exercise will likely suffer. Such incentives and capacities would be products of the overall policy and institutional frameworks for resource management in the federal administration. How the PPA model as a whole is designed and managed plays an important part of this larger puzzle. Thus, it is our contention that improving the effectiveness of the PPA evaluations, no matter what the methodological approach adopted – and here there clearly are different options for improving the methodological foundations of the annual evaluations – depends ultimately on the effectiveness of the PPA model as a whole.

5. OPTIONS FOR THE PPA AND EVALUATION

Making Sense of the Limited Effectiveness of the PPA

69. We have so far documented findings of our rapid assessment of the effectiveness of the PPA model and its annual evaluation system. Although our methodology is not empirically rigorous, we believe the work covers reasonable ground to draw inferences about the effectiveness of the PPA and its evaluation system. Assuming our findings are not somehow skewed representations of a biased minority (i.e., the majority of officials whom we did not interview hold a far more positive views of concrete contributions of the PPA), it is then necessary to develop an set of explanations as to why the PPA has failed to live up to its original expectations. Based on a review of international and Brazilian literatures about government planning and management as well as our own intuition and interviews, we have come up with three broad types of possible reasons why the PPA has proven ineffective. These are divided into (i) problems that have arisen from the initial design of the PPA model; (ii) external constraints; and (iii) problems related to implementation and management of the PPA process. A more detailed discussion is found in the Appendix 4, which separates problems arising from initial implementation of the PPA from those related to the current management of the PPA. Here we have collapsed these two categories into one set of hypotheses.

Box 4: Hypotheses about the ineffectiveness of the PPA

Initial PPA design problems:

- The program format is a problematic organizing principle for planning and management.
- The program format is incompatible with other organizing methods in the agencies.
- The management model is too simple and standardized to serve as a management tool useful for the agencies.
- The management model is too complex to serve as an accountability mechanism between the center (SPI) and the agencies.

Problems of the PPA's external environment:

- The PPA is undermined because it is not truly integrated with budget making and execution.
- The agencies can face political incentives and structural problems incompatible with the PPA.
- The PPA does not receive the political support of this government.

PPA implementation problems:

- The SPI did not make an effort to get the agencies to 'own' the PPA.
- PPA Programs were misidentified because the methodology really meant that problems were identified ex post from the existing set of actions that were weaved together as programs.
- The SPI gave too few resources to the agencies to prepare for the PPA.
- The SPI is missing opportunities to help the agencies.
- The SPI is missing opportunities to help the center of government.

70. The crudeness of the methodology used for our assessment does not allow us to discriminate among these hypotheses. We have found evidence that is consistent with every one of the hypotheses. Our judgment is that all three sets of causes have contributed to undermining the effectiveness of the PPA.

Initial design flaws

71. The passage of time suggests two fundamental weaknesses of the PPA model, the insistence on a “pure” program format and a choice of universality over selectivity. The ambition of the “pure” approach was to put *all* government actions within a standard program format *and* to make programs the central logic for organizations. The idea makes intuitive sense, and the program format has undoubted virtues such as transparency and an emphasis on outputs/outcomes. By the standards of other reforming countries, this was a high ambition indeed. But programs cannot always be unambiguously identified, while the program logic may be at odds with organizational logic and with budget incentives – in particular, this logic may have less sense for routine than for developmental activities.

72. Universality ignores differences between organizations in terms of their objectives and technical capacities; assumes that better performance will come about through a standard formula (i.e. “one size fits all”), rather than one that differentiates between objectives and between capacities; spreads resources too thin; fails to signal true priorities of the government (and thus also lead to less support from the Presidency itself); and creates bureaucracy without results. Therefore, we now believe that trying to apply a standardized approach to improving management in different types of activity prove ineffectual.

73. It is therefore not surprising that the PPA, in particular the program format, has come to face substantial resistance. Sectors have resisted this format in order to preserve budget flexibility (e.g., by trying to define as a program an umbrella set of actions, thus rendering the “program” mere unit of budget classification and appropriation rather than a unit of management) and/or to preserve their existing organizational structures, whether out of conservatism or because the program logic does not work as organization logic.

74. *“Pure” program concept:* The decision to apply the same logic of program design – as logical as the problem-driven approach may have seemed – to different types of government activities and use it as the basis for structuring the budget could have led to misfit in some cases. The question here is whether the problem-oriented definition of programs is always appropriate. Broadly speaking, different PPA finalistic programs cover routine activities (maintaining roads, making social-security payments), developmental activities (implementing a new policy such as decentralization of health services, or developing a regional pole), crisis activities (such as handling outbreaks of diseases), and policy-making activities.²² Of course, all activities can be improved, but the problem approach (like its close cousin the logical framework) may be more appropriate, for organizational purposes at least, for developmental activities than for the other three types.

75. *Universality:* We understand the logic behind the universal approach at the very beginning of the model’s launch in 1999, but this also relates to the government’s inability to set sharply defined policy priorities. Whatever the original reasoning, it is our

²² These four categories are suggested in Handy (1993).

view that the universalization limited the range of options available for managing the PPA implementation (e.g., no room for piloting) and pushed the SPI in the unenviable position of being in the middle of detractors and potential supporters of the PPA model. On the one hand, the model has been “rejected” by those who did not find the PPA approach a natural good fit to their way of doing business or to their own aspiration about how to improve performance. At the same time the universal coverage of the PPA forced SPI to stretch its limited resources across the board, which may have prevented it from providing much needed support to those ministries that have shown greater willingness to use the PPA as a tool for their own managerial transformation.

76. Besides affecting the SPI’s strategy managing the PPA implementation, the universal application has invited symptoms of bureaucratization rather than nurturing development of performance orientation in ministries in tailor-made ways. The only feasible way for managing such a complex system for the entire government was to develop a series of routines, but ironically, these routines returned the PPA to the bureaucratic manner of managing resources and away from the theoretical ideal of managing by results (which would have required less standardized and more case-by-case entrepreneurial approach to problem-solving).

77. *Plan-budget linkage:* A natural consequence of the decision to adopt programs as the basic unit for organizing government activities and to do so across the board was to restructure the budget on the basis of these programs. Linking the PPA to the budget (or rather making the two basically identical in form) was also a response to the typical problems of government planning, also observed in Brazil in prior years, where plans did not necessarily guide resource allocations because they were divorced from the annual budgeting exercises. This has created some benefits as well as perverse incentives.

78. On the benefit side, the PPA’s program format and its adoption as the structure of budget appropriation have significantly increased fiscal transparency. Now oversight bodies such as Congress and the TCU can monitor budget execution by program, and reasonably detailed qualitative information about each program (beyond its title) is available for those who are interested in understanding what each of them does. The TCU has begun to conduct its annual audits by program, and started a credible program of selective program evaluations (although the TCU evaluations do not limit themselves to PPA programs as the unit of analysis). Among the most positive consequences we came across during our interviews were the cases where the ministries were mandated to improve their program and overall management by the TCU after its audit found deficiencies in the ministries’ program management.

79. On the negative side, however, the use of the PPA program classification as a structure of budget appropriation gives ministries incentives to aggregate programs into larger “umbrella” programs so as to retain greater flexibility during budget execution. This in fact seems to be what happened in those large social sector ministries. Greater flexibility in budget execution makes sense from the point of view of the executing agencies, and is the general trend in performance-based budgeting among OECD countries. But, it can have a detrimental effect on the PPA to the extent it removes one of the few tangible benefits it has brought about, transparency. We also imagine that this

sort of “opportunistic gaming” where SOF/SPI and the line ministries negotiate program definition with different motives – SPI wanting to preserve the conceptual purity of what a program should be, SOF trying to limit line ministries’ spending increase, and line ministries trying to lobby for more budget – would be unhealthy for promoting performance orientations within the government. In the least, it would foment cynical attitudes about programs among line ministries.

80. It can be said that one instrument cannot satisfy two objectives. One instrument, the PPA program, has difficulty in serving the multiple objectives of planning/ policy making (in the sense of identifying problems-programs and rationalizing the inputs associated with them) and resource and organizational management. For instance, the US Department of Defense’s program budget is used for planning, but appropriations are made by organization (see Appendix 1 for details). Had the PPA’s program classification not been the basis for budget allocation, it might have corresponded to the Department of Defense’s model of policy rationalization. But it was made the basis for budget allocation, which led to budget games – the manipulation of programs by both sides to gain control of budget resources. Kim et al (2006) take it as an axiom that two different functions related to the budget, such as planning and spending, be kept separate: “it is exceedingly difficult to subordinate the organization structure, even when the government has a program budget, because organizations actually spend the money and are responsible for results.” (page 27).

External constraints

81. Aside from the inherent limitations in the original design of the “model,” the lack of political support has been harmful to the further development of the PPA under the current administration. The PPA’s loss of political support is partly the result of bad luck, partly a consequence of PPA-specific failures. The PPA lost support under the new Administration in 2003, partly because of the incoming government’s distrust of the bureaucracy and different priorities by the new minister of planning who appeared to be more interested in economic development issues via industrial policy, and later on, via development of a new law for public-private partnerships than public management reforms. One might also infer that the lack of political support arose from the PPA’s inability to deliver. This resulted in the transfer of initiatives to other parts of the government – monitoring of key programs to Casa Civil, the leadership role in the quality of public spending agenda to the National Treasury, strategic thinking to “Brasil em 3 Tempos”, and so on. The failing credibility of the PPA has also meant a loss of support from the ministries, the Ministry of Finance, and SOF.

82. The same political context that led to the dwindling political support for the PPA also explains the lack of further development in complementary policies that were needed to support and strengthen the PPA. The most important failure in this regard has been in the area of expenditure management where the combination of extremely high levels of earmarking and policy emphasis on fiscal balance have virtually eliminated the PPA’s capacity to influence budget prioritization. A government-wide plan without the ability to establish and protect policy and spending priorities is a handicapped instrument from its inception. Other well-known external constraints include the rigidity in human

resource and materials management (including the highly bureaucratic procurement process), many of which have been repeatedly identified in the annual evaluations and the periodic assessment of the plan management (Box 3).

83. Effectiveness and efficiency of program implementation also depend on other public management functions such as human resource management and government procurement. In the former, the government has recently undertaken no major initiative to make human resource management more performance-oriented. In procurement, the government has adopted some promising measures such as an expanded use of *pregão eletrônico*, and rationalization of government purchases (at least in the first year or two). But the cumbersome process involved in government procurement remains a constant complaint of program managers, and a major overhaul seems beyond the horizon.

Management of the PPA implementation

84. The Ministry of Planning in general, and the SPI in particular, may have missed opportunities, on several fronts. The SPI has not been able to act, apparently for lack of appropriate authority and/or resources, to respond to implementation difficulties faced by programs and sectors. Similarly, the MP (primarily SPI and SEGES) has not been able to put to use the management information collected during the evaluation process. For example, the SPI was not able to respond to the government's information needs associated with PPI. Had it been a center of knowledge and of strategic-thinking, it could have informed the government about the poor state of readiness of the initial set of PPI projects. The PPA's ability to protect resource allocations to priority programs, weak to begin with, has weakened further with the decision to abandon the cash-flow control. SPI could have done more to publicize the PPA or evaluation results or to get feedback about the PPA from the ministries, although we would never know whether such a pro-active "marketing" would have been sufficient to maintain a needed level of government support for the PPA.

85. In our view, SPI has missed the opportunity – because it chose the road of universality rather than selectivity – to improve specific elements of the sectors' capacity for performance management, notably in the areas of information, management (and indicators), evaluation capacity, and planning capacity. In hindsight, it seems to us that SPI has tried to create a system before the sectors had sufficient capacity for this, and in so doing, neglected to pursue an alternative path of building on different ministries' strengths and weaknesses in a more tailored approach that would leverage ministries' own interests and commitments to a higher degree.

86. The bottom line, then, is that the effectiveness of the PPA has been weakened from the beginning by certain design choices made by the SPI and further undermined by both the lack of buy-in from the line ministries and by the hostile external environment. There is probably little that the SPI could have done to deal with the hostile external environment itself. But then, a fuller recognition of what this meant for the potential of the PPA might have led to a more modest design and approach to managing the PPA process and a more realistic tactic of managing ministries' expectations.

87. We also believe that there must have been alternative ways of managing the relationships with the ministries so that they, or at least some of them, would develop greater ownership of the PPA. This certainly would have required abandoning a relatively rigid universalist approach, and adoption of a ministry-by-ministry approach of collaborative problem-solving.

88. Based on these interpretations briefly sketched above, we will offer some suggestions for how the PPA concept could be revamped to pursue its original objective – to make the federal administration explicitly results-oriented.

Reviving the PPA Model

89. We have suggested that the model suffers both from its universal (rather than selective) application and from its “pure” program form and that this combination has made it rather bureaucratic than problem-oriented. From this observation, we draw the conclusion that one way to revive the PPA is to abandon these two characteristics, and make the PPA more selective (and thus strategic) and pragmatic and less controlling (i.e., let those ministries ready for performance management catalyze on what the PPA has to offer rather than impose the model on all irrespective of their interests and preparedness). These would in turn have clear implications for how to structure the evaluation policy associated with the PPA.

A selective PPA

90. To make the PPA more selective, it makes sense to consider returning in the direction of the 1996-1999 model of the PPA by restricting the strategic focus of PPA to priority programs where focused attention of central agencies such as the MPOG is likely to bring about positive pressure for improved performance without overstressing its capacity for managing the process. One way in which the MPOG could provide focused attention is through preferential treatment of priority programs in budget prioritization. Thus, those ministries which have a strategic role, yet come at the bottom of the queue for budgetary resources (e.g., infrastructure ministries) would constitute possible candidates for such treatment.

91. The PPA could also serve as a platform of performance management in those ministries that have their own impetus for performance orientations, and are looking to central agencies such as SPI and SEGES for support. In our interviews, the Ministry of Agriculture came across as one such ministry. At least this ministry’s planning area appeared strongly interested in augmenting its capacity, and was already seeking technical support (e.g., training) from MPOG. Another example of good management was the Ministry of Tourism. Although this ministry’s managerial dynamism apparently owes to the entrepreneurial attitude of the minister, working closely with a ministry like this could have been an interesting tactical choice for SPI to develop positive demonstration cases. Table 4 suggests a three-way categorization of ministries:

- High priority for the PPA: ministries with strategic importance, but without protected budgetary resources (and typically weak managerial capacities).

- Middling priority for the PPA: smaller, “non-strategic” ministries that may benefit from PPA support because of their commitment to performance improvements and openness to work with the Ministry of Planning to this end.
- Low priority for the PPA: ministries with large and protected budgetary resources which are relatively resistant to the influence of the PPA (although a number of their programs may be official government priorities, their designation as such in the PPA and focused attention from MPOG are unlikely to add much value to their performance because of the size and the political strengths of these ministries and the relatively assured funding they enjoy).

Table 4: A Proposed Typology of Ministries for a Selective PPA

Type of ministry*	Ministries	SPI Role
Ministries under heavy political interference, perhaps because they have visible capital investment projects. They may or may not have an institutional history. (SOF: “The infrastructure ministries have their own agenda. If they have better planning capacity, they’ll use it for that agenda.”)	Cities, Transport (in 2006)	Politics prevents SPI getting involved effectively, but these are sectors where planning is important, so a Brasil-em-Ação-type selective mechanisms could be revived to concentrate on certain strategic programs/actions (and leave out the rest).
New [or small?] ministries looking for a model. They may not be excessively politicized.	Environment (2001), Tourism, Culture, Agriculture, Science & Technology	This may be the most fertile ground for SPI technical assistance, but these ministries typically do not constitute priorities or account for large spending. SPI should help these ministries on request only.
Well-established (“old”) ministries with finalistic programs, often with large and protected budgetary resources and (for better or worse) their own management model. (Two of these ministries, MEC and MS, have large sub-national responsibilities and their front-line professionals provide a counter-weight to pork-barrel politics to some extent.)	Health, Education, Social Development, Labor and Employment	The center of government does not have a budget lever, so SPI should broadly leave these ministries to manage themselves (unless they ask for help) and limit its mandate to requiring annual reporting based on a pre-specified format (e.g., based on the current model of annual evaluations).

- Criteria: degree of budget protection; degree of politicization wsr budget amendments; age

92. For high-priority ministries, the central idea would be for MPOG to provide support, along the lines of the 1996-99 PPA model, in exchange for some form of budgetary preferment based on management reforms and/or good performance. But this tutelage should be only for some of their programs (or projects). It would be necessary to apply an SPI-controlled budget-preferment mechanism (like cash-flow control or the PPI mechanism) for this, as is currently done with the PPI projects.

93. One way to make this arrangement concrete would be for SPI, together with the relevant line ministries, to prepare a “chapter” of the PPA focused on government-wide strategic priorities, similar to *Brasil em Ação* in the PPA 1996-99. This would serve as a “compact” between MPOG (as representative of the government) and each of the participating line ministries and give SPI co-responsibility to oversee the effectiveness of this part of the overall plan (i.e., through selective monitoring) and support its implementation (e.g., via a mechanism for budget prioritization such as the cash flow

control). Assuming the government's current prioritization of the infrastructure agenda continues, an obvious choice would be for MPOG and the infrastructure ministries to develop a government-wide medium-term infrastructure plan (strategy) as a basis for this special "chapter" of the PPA.

94. Likewise, the SPI (and the SEGES) should become a source of expertise for these ministries, in particular by helping to develop these ministries' capacities for strategic management. MPOG would become a repository of strategic information about these ministries; and it might even consider merging SPI and SEGES under a single command (along the lines of SEPLAG of Minas Gerais, for example) and turn it into a promoter of strategic management change in the federal administration: planning, along with other organizational and managerial instruments, would become one of the array of instruments MPOG would deploy to make these selected ministries more performance-oriented.

95. This tutelage role would change MPOG's role in several ways, specifically to:

- build on its existing sectoral expertise, but focus on those priority sectors in which the strategic planning capacities within the ministries are still weak – this would mean SPI would have to re-acquire skills for policy analysis necessary in the infrastructure sector; and
- become a consultant in strategic management especially for those ministries in the middle category above that demonstrate clear commitment to enhancing their performance orientations with MPOG support (which would have implications for getting the resources of SPI and SEGES to work together).

96. What would happen with non-priority ministries (the third category above)? SIGPlan reporting requirements should be eliminated or minimized. We do not see much value-added in having a central agency like SPI "monitor" the entirety of government programs, especially when it cannot act on the information to alleviate constraints that are identified through the system. However, as the principal administrator of the whole government apparatus and its budget resources, it would be perfectly legitimate for MPOG to demand accountability from line ministries. We believe, however, that this should not take the form of current SIGPlan with the expectation of "real-time" data entry and central monitoring. Each of the line ministries should have a robust monitoring system of its own, but the day-to-day information needs at the level of individual ministries could not conceivably be met with a central instrument like SIGPlan. Instead, MPOG should demand accountability from line ministries on the basis of periodic self-reporting, probably on an annual basis, or at most, every semester.

97. The current annual evaluation exercise would serve as a good basis, although SPI might consider treating programs differently depending on their importance vis-à-vis the government's overall policy priorities. In any case, the periodicity of this reporting requirement should probably not be more frequent than twice a year to avoid overloading line ministries. The government might also consider the option of mandating each ministry individually to report on its performance directly to Congress and discontinue

the current process whereby the SPI prepares voluminous tomes of annual evaluation reports that consolidate ministry evaluations.

The program format

98. Having invested in creating a program format, with its benefits of transparency, SPI should maintain it. But it should probably give the ministries greater leeway to decide on the definition of programs (albeit with SPI advice). What matters more than the conceptual purity and its application across the board is a good fit between each ministry's overall objective, its existing managerial and internal accountability structure, their policy objectives, and the definition of programs. As argued above, a number of the government activities are not easily defined and organized as programs because of a variety of confounding factors such as the multiplicity of objectives, inter-connectedness and inter-dependence with other activities, difficulty of defining a clear, monitorable goals, etc. Inevitably, program design is more an art than a science, and it should be the ministries' responsibility to exercise this art.

99. Transparency and accountability should always be demanded, even of these more complex types of government activities. But it probably would not benefit the PPA and MPOG to insist on a blanket application of a uniform set of criteria for program definition across the board because that would generate resistance, skepticism, and resentment. This might mean that not all the actions in a given ministry need to be organized as programs. At the same time, however, those ministries that are less able to design and present their programs in a rational fashion should be subjected to greater MPOG scrutiny during annual budget negotiations (i.e., those who can't explain what the resources are for should not receive funding as easily as those who can).

100. Giving ministries greater leeway does not mean anything goes. Certain parameters should be established to maintain the integrity of the whole plan. For example, initially at least, the MPOG may want to maintain a control over the size of programs (and seek to avoid large umbrella programs) if the government is not yet prepared to allow ministries greater control over their budgets. It would also want to control the total number of programs so as to avoid atomization of the plan. This particular suggestion, however, would become null if our separate suggestion to "de-link" the PPA from the budget appropriation structure were accepted (as suggested below as a possibility).

Complementary Policies: Budget Reform

101. Several aspects of the current practices in budget management tend to undermine the PPA's capacity to pursue strategic priorities and to promote performance-based management. The current expenditure composition is not conducive to promoting economic growth, and there is much inefficiency in execution of the public resources.²³ Ideally, the PPA and a broader reform of the planning and budgeting system should

²³ For recent Bank analyses of these issues, see *Brazil: Improving Fiscal Circumstances for Growth*, Report No. 36595-BR (June 29, 2006) and *Brazil More Efficiency for Better Quality: Resource Management in Brazil's Unified Health System (SUS)*, Report No. 36601-BR (June 29, 2006).

progress hand-in-hand and complement each other. At the moment, however, it seems that the PPA's effectiveness is hampered by the complexity and the perverse incentives the existing planning and budgeting system generates, while the PPA's contribution to improving the rationality of planning and budgeting is limited.

102. Thus a critical question is how to unleash a virtual circle in which a broader reform to the planning and budgeting system provides a positive enabling environment for the PPA, while the latter makes concrete contributions to strengthening federal public expenditure management. This, in our view, would involve clearer assignment of roles to a set of existing instruments that seem to have unnecessarily overlapping functions, and to do so with a clear view of how to overcome substantive weaknesses in public finance management, such as under-investment in infrastructure, inefficiency and limited accountability in social service delivery.

103. Earlier we reported on our finding that some ministries (at least for some of their programs) seem to approach program design not from the point of view of performance enhancement (i.e., via clearer specifications of the program's objectives, logical framework, and input requirements), but with a view to increasing flexibility during execution. This is accomplished by defining programs as broadly as possible so that more of the ministry budget can be re-allocated during execution without external approval. It is perfectly reasonable for line ministries to want greater flexibility, although this needs to be controlled by the central budget authority to make sure there is adequate ex ante control. What is unfortunate is that in the middle of this budget "game," the expected benefits of the PPA's program approach end up being sacrificed.

104. One radical solution would be to reverse one of the innovations of the PPA 2000-03, the adoption of the program classification as the structure for budget appropriation. This was done to avoid the historical problem (which is also common outside Brazil) of having a plan that is divorced from the actual resource allocations (via annual budgets). We still believe the objective of integrating plans and budgets remains valid and important. But by insisting on a particular form (i.e., the identical structure for the plan and the budget at the level of individual programs and actions), the integration has been achieved in form, but not necessarily in substance. It is our view that the plan (PPA) should indicate government priorities and provide a framework of accountability for the whole government. This, however, can be accomplished without the identical structure between the plan and the budget. To the extent the use of the identical structure invites the kind of "gaming" discussed above, it would seem preferable to adopt different structures for the plan and for budget appropriations.

105. For example, the plan should still be based on the program structure similar to the current format, except it may no longer be comprehensive (given our proposal to make the PPA more selective) or may no longer contain the same details at the level of actions. The budget appropriations in turn could be done on the basis of the ministries' organizational structures – after all it is not the programs but the organizations that spend the money and are held accountable for results. There should be certain ex ante control over how the appropriated funds can be used (e.g., restriction for using the capital budget to finance recurrent activities). The level of appropriations could be at the level of

secretariats and equivalent organizational units in the indirect administration, or possibly at the level of coordinations. These units, in turn, should base their own budget proposals on their own annual plans regarding implementation of programs and other activities. In this modified structure, SPI could, and probably should, continue to press the ministries to design programs in as conceptually pure a way as possible, avoiding umbrella programs that do not mean much in linking resources assigned and results pursued. SOF would determine budget allocations to these units on the basis of their overall performance in the previous budget year (and before, when relevant) as well as the credibility of each unit's budget proposal (which will be aggregated to form a ministry's budget proposal). An array of tools including program evaluations and expenditure reviews can be institutionalized to support budget decision-making within the Executive (and ideally by Congress as well).

The Evaluation Model

106. Mechanisms to provide evidence-based information to aid decision-making can contribute to improving planning and budgeting. By the same token, program evaluations can play a useful role, and this is one of the areas in which MPOG is currently investing a considerable amount of efforts. As important as it is, we also contend that program evaluation is a secondary tool for strengthening managing of public policies and resources. Good program evaluations in dysfunctional budget management environment will probably do little to better the quality of government decisions. Thus making marginal changes to the existing model will be unlikely to make the system viable and certainly not address the core problems of the PPA. The current practice of self-evaluation could evolve into annual performance reporting by the ministries, as suggested above. But such a change is still unlikely to provide MPOG with the kinds of information needed to make informed decisions about expenditure allocations.

Annual evaluations for a selective PPA

107. If our suggestion to adopt a selective PPA is acceptable to the government, then that would lead to specific implications for how to design the PPA evaluation process. First of all, if a small set of programs/projects are identified as government-wide priorities, then it is logical that these programs/projects should be monitored and evaluated with the participation of the central organ such as the MPOG. For other programs, the responsibility for monitoring and evaluation should be fully decentralized to the ministries, possibly under a minimum set of requirements such as the example of Australia's multi-year departmental evaluation plans under which the departments were required to evaluate all of their "major" programs at least once in a cycle of several years. Thus the MPOG could issue a new evaluation policy that mandates each ministry to evaluate within the PPA cycle a set of programs or actions based on pre-established criteria and report the findings directly to Congress.

108. In both cases (i.e., government-wide priority programs to be evaluated jointly by the ministries concerned and the MPOG and those programs to be evaluated solely by the ministries), how each program should be evaluated would be up to the characteristics of the program and the types of priority issues facing them. The MPOG might still issue

technical guidelines, but not as mandates but as references that the ministries are free to follow. If the ministries lack own capacities to design technically sound evaluations, they might seek MPOG's technical assistance, for which the Ministry should develop adequate technical preparedness.

109. A key consideration in determining the evaluation method to be adopted is the possibility of utilization once the findings are made available. For this, it is important to keep proper alignment among (i) who will lead the evaluation process – whether this is done in-house or contracted-out; (ii) what information the evaluation is expected to provide; and (iii) who has the authority and capacity to act on the findings. In our view, the lack of this alignment has been one of the reasons for poor utilization of the PPA annual evaluations. For this reason, those evaluations led by the ministries should focus on identifying deficiencies that the ministries themselves can correct, such as program design or implementation difficulties arising from internal management. In contrast, the evaluations of those government-wide priorities which the MPOG would participate in should include identification of specific external constraints, provided that the MPOG is empowered to deal with at least some of them.

Annual evaluations and budget formulation

110. In adjusting the approach to annual evaluations, one option the MPOG might consider is to forego any intent to use program evaluations as inputs for budget decision-making, at least for the time being. Links between evaluations and budgetary decision-making are nebulous in most places, and successful cases of systematic use of evaluation information for budgeting are very limited in number (Box 9). The Canadian model of evaluation which inspired part of the latest innovations by the MPOG is itself focused more on internal management and accountability at the department level rather than budget allocations. Once evaluation results are available, neither the MPOG nor the sector ministry would or should be prohibited from using the information during budget negotiations. But the rather over-engineered process of annual revisions of all the programs based on annual evaluations of all the programs should be abandoned because of the high transaction cost with dubious benefits in terms of improved quality in program design and better efficiency in budget resource allocations.

111. A more efficient approach would be to emulate the Australian model of selective review of new programs and projects for additional funding each year. When these relate to a request for continuation or expansion of an ongoing program, then it would be reasonable for the MPOG to expect robust program evaluation as part of the background materials the ministry can provide to justify additional funding. This alternative would thus be consistent with the current shift in emphasis toward selective program evaluations. However, for this model to be fully workable, the government will need to develop a capacity to estimate expected levels of additional resource envelope for funding new initiatives on an annual basis.

Box 9: International Best Practice in Evaluation

There are (a few) alternative models for a government-wide evaluation system, but their character is substantially different from the PPA's system. Very few countries have successfully run government-wide evaluation systems. The number does not extend beyond Australia and Chile (and perhaps Canada). Their systems are far more selective, with methodologies that allow sector-specific analysis. They use expert evaluation (though sometimes sectors evaluate themselves). They are consistent with incentives that encourage the sectors to act on evaluation findings.

Chile presents a model of a centralized evaluation system that works, but evaluations are highly selective (50 a year), they use a logical-framework methodology for sector-specific analysis, they are carried out by external experts, and the central agency has the power and the expertise to impose and manage *some* change at the sector level. Under present circumstances, the authority at the center of government in Brazil (let alone SPI by itself) is not strong enough to make a centralized evaluation model a la Chile a feasible option.

Source: Appendix 1 of Shepherd (2004).

References

Ataide, Pedro Antonio Bertone (2005), *Avaliação do Plano Plurianual: análise das restrições à sua integração ao ciclo de gestão pública federal*, Dissertação (mestrado em administração) - Universidade de Brasília, Programa de Pós-Graduação em Administração (<http://www.unb.br/face/ppga/arquivos/dissertacoes/Avaliacao%20do%20Plano%20Plurianual.pdf>)

Brasil (2005) *Manual de Elaboração de Programas Plano Plurianual 2004-2007*, Brasília: Ministério do Planejamento.

Calmon, Katya Maria Nasiaseni e Gusso, Divonzir Arthur. *A Experiência de Avaliação do Plano Plurianual (PPA) do Governo Federal no Brasil*. Brasília: Planejamento e Políticas Públicas, nº 25, jun/dez 2002; Instituto de Pesquisa Econômica Aplicada (IPEA).

Coutinho Garcia, Ronaldo (2000), *A Reorganização do Processo de Planejamento do Governo Federal: o PPA 2000-2003*, Texto para Discussão nº 726. Brasília: IPEA

Divorski, Stanley (1998), 'Evaluation in the Federal Government of Canada', In Mackay, K. (ed) *Public Sector Performance – the Critical Role of Evaluation: Selected Proceedings from a World Bank Seminar*, Washington, DC: The World Bank

Furubo, J.E. (1994), "Learning from Evaluations: The Swedish Experience," in F.Leeuw, R.Rist, and R.Sonnichsen (eds.), *Can Government Learn? Comparative Perspectives on Evaluation & Organizational Learning*. New Brunswick,NJ: Transaction Publishers.

Gaetani, Francisco (2003), "O Recorrente Apelo das Reformas Gerenciais: uma Breve Comparação", *Revista do Serviço Público*, Ano 54, n. 4, 106p., Out-Dez

Garcia, Paulo Francisco Britto (2002), *A "Procustomania" na Elaboração e Gestão do PPA 2000-2003: a Prática Determinista Inconsciente Preside a Formulação do Plano: o Planejamento Estratégico Situacional como Ferramenta de Governo*, Dissertação apresentada à Escola Brasileira de Administração Pública da Fundação Getúlio Vargas, como exigência parcial para a obtenção do Título de Mestre em Gestão Empresarial Pública, Brasília

Garcia, Ronaldo Coutinho (2001), *Subsídios para Organizar Avaliações da Ação Governamental*, Instituto de Pesquisa Econômica Aplicada (IPEA), Texto Para Discussão No 776, Brasília, janeiro de 2001

Giacomoni, James (1997), *Orçamento Público*, São Paulo: Editora Atlas.

Jenkins, B. and A. Gray (1990). Policy Evaluation in British Government: From Idealism to Realism. In R. C. Rist ed. *Program Evaluation and the Management of Government*. New Brunswick, NJ: Transaction Publishers. 53-70.

Kim, Dong Yeon, William Dorotinsky, Feridoun Sarraf, and Allen Schick (2006), "Paths Towards Successful Introduction of Program Budgeting in Korea", Chapter 2 of John M. Kim., ed., *From Line-item to Program Budgeting: Global Lessons and the Korean Case*, Korea Institute of Public Finance

Mackay, Keith (2003), *Two Generations of Performance Evaluation and Management System in Australia*, OED. Available at www.worldbank.org/oed/eed/

McDonald, Robert, (2002), Practical Approaches to Support Evaluation in the Government of Canada, European Evaluation Society Conference 2002, Sevilla, Spain, October.

Mayne, J. (1994) 'Utilizing Evaluation in Organizations: The Balancing Act', in F. Leeuw, R. Rist and C. Sonnichsen (eds) *Can Governments Learn? Comparative Perspectives on Evaluation and Organizational Learning*, pp. 17–43. New Brunswick, NJ: Transaction.

Mintzberg, Henry, Bruce Ahlstrand, and Joseph Lampel (2000), *Strategy Safari: A Guided Tour Through the Wilds of Strategic Management*, New York: Free Press

OECD (2003), *The Learning Government: Introduction and Draft Results of the Survey of Knowledge Management Practices in Ministries/Departments/Agencies of Central Government*, GOV/PUMA (2003)1 ([http://appli1.oecd.org/olis/2003doc.nsf/linkto/gov-puma\(2003\)1](http://appli1.oecd.org/olis/2003doc.nsf/linkto/gov-puma(2003)1))

Schick, Allen (2001) *Does Budgeting Have a Future?* OECD (PUMA/SBO (2001) 4). Paris:OECD.

_____.(2003) *The Performing State: Reflection on an Idea Whose Time Has Come But Whose Implementation Has Not*. OECD (GOV/PUMA/SBO (2003) 17). Paris: OECD.

Wilson, I (1994), "Strategic Planning Isn't Dead— It Changed." *Long Range Planning*, 27, 4, August, pp.12–24

World Bank (1998) *Public Expenditure Management Handbook*, Washington: World Bank.

World Bank (2002), *Brazil Planning for Performance in the Federal Government: Review of the Pluriannual Planning*, Report No. 22870-BR, Washington: World Bank.

World Bank, (2002) *Annual Report on Evaluation Capacity Development*. OED. Washington: World Bank. Available at www.worldbank.org/oed/eed/

World Bank (2006), *Chile: Study of Evaluation Program, Impact Evaluations and Evaluations of Government Programs*, Report No.34589- CL, Washington: World Bank.

APPENDIX 1: THE EXPERIENCE OF PROGRAM BUDGETING IN OTHER COUNTRIES

1. A chapter of a recent study on introducing program budgeting in Korea reviews the international experience (which, by inference, does not appear to be very extensive).
2. Making program budgeting work is very difficult. There have been many failed designs and failed attempts to implement it. A major difficulty in implementing program budgeting is to deal with the *multiplicity of government purposes*. For instance, do schools teach children to be productive or to be good citizens? do health clinics in schools serve education or health needs?
3. Program budgeting can have different possible objectives, and specific objectives fit specific country situations. Program budgeting is a tool for four different functions: policy analysis; improving government performance; accounting for the full cost of government activities; and planning for future spending options.
4. ***Program budgeting as a tool of policy analysis.*** Program budgeting can help in evaluating the cost-effectiveness of alternative spending options that have the same objective. Ideally, for policy analysis, programs that did not observe organizational barriers would be used to allocate resources while appropriations would be made, based on the programs, by organization. In practice, program budgets have been run on a hybrid model that combines programs and organizations within a single structure. In the US Department of Defense, program budgeting as a tool of policy analysis grew out of the wartime experience of allocating resources to buy war materials and choosing between alternative ways to pursue the same objective. Analysis based on programs has proven effective in re-evaluating, and changing, policies.

“Each program should have a single, identifiable (and preferably measurable) end purpose that is distinct from the activities that government is carrying on. In other words, programs should be defined independently of what government is doing. As elementary as this step seems to be, it has often been among the most difficult and controversial step in program budgeting. This form of reasoning led to a significant realignment of U.S. defense forces after World War II. The Air Force operated a fleet of long-range bombers whose stated purpose was to penetrate enemy defenses so as to prevent an attack on the United States. Policy analysts, however, defined the purpose as surviving an enemy attack with the capacity to launch a counterattack. The Air Force significantly redeployed its forces in response to this revised definition of end-objectives.” (page 28)

5. The State of Hawaii tried a very ambitious form of pure program budgeting, with full costing and covering a six-year cycle. Programs were defined at nested levels, 11 State programs, 340 intermediate programs, and 580 at the lowest level. Information requirements from dealing with 580 programs and full cost accounting overloaded the system. Legislators were unable to absorb the details. Hawaii had to amend the law to simplify things.

6. ***Program budgeting to improve government performance.*** This later version of program budgeting reflected the philosophy of the New Public Management: output budgeting was intended to free managers of the shackles of traditional input budgeting and controls. This approach also changed the nature of budget discussions at the center of governments: inputs were no longer the focus, outputs and policies were. This version of program budgeting is less insistent than the first on the primacy of programs over organizations: at lower levels, organizational units are allowed to be programs.

7. Australia's financial reforms in 1984 were intended to give focus on results and give them greater spending discretion to do this. A program classification was used for expressing targets and for reporting, but the organizational structure continued to be used for policy decisions and appropriations. This program structure generated large amounts of information that was not used by government or parliament. In the late 1980s the system was replaced by a new system that would make results budgeting effective. Ministry responsibilities were redefined to identify 17 broad portfolios in 17 ministries. Each ministry had broad spending discretion within its portfolio (which effectively replaced the programs.). "Paradoxically, although the program structure became largely irrelevant, the goals of program budgeting were largely achieved. Budgeting pays much more attention to results, program evaluation has been applied more extensively and systematically than in any other country and performance data are published in the budget. But Australia has done all these things within the portfolio structure, which is organized along departmental lines. Perhaps this fate is inevitable, for once the Government opted to cast reform in managerial terms, it was impelled to subordinate programs to organizations." (page 33)

8. ***Program budgeting as a means of accounting for the full cost of government activities.*** Both the policy-analysis and government-performance versions of program budgeting need, ideally at least, programs to be fully costed, but the former wants this on the basis of programs, the latter on the basis of organizations. Full costing is difficult, and this gets worse the greater the number of programs.

9. New Zealand has progressed much further than other countries in using accounting as a management tool. The budgeting system it adopted in the late 1980s is similar to managerial versions of program budgeting. The systems permitted budget appropriations to be made on the basis of outputs. But because of the difficulties in developing cost accounting and allocation systems, most appropriations are still based on input costs.

10. ***Program budgeting to enable government to plan ahead and set spending options.*** In many countries, a medium-term expenditure framework (MTEF) has replaced national planning as the main tool for taking government decisions about the future. National planning coordinates the public and private sectors, while an MTEF deals only with the public sector. In most countries which still prepare national plans, the annual budget "has the last word, leaving the plan as a string of unfulfilled promises." For countries with national plans, program budgeting can make budgeting a more strategic exercise and make planning more cognizant of fiscal constraints.

11. Brazil is one of the few countries to make a sustained effort to integrate budgeting and planning. Despite the problems it faces of integrating program management with the existing organizational structure, “Brazil has shown that it is feasible to plan under a fiscal constraint, and to budget in a more strategic manner.” (page 37)

Some Comments on the Experience

12. The chapter described above discussed five cases of implementing a program budget. The reader cannot infer success in all these cases, but the cases of the US Department of Defense and Australia, both successful, stand out as relevant to Brazil. This is because the PPA set out to be an instrument both for better policies (the problem-program approach) and for better management. There appear to be two principal lessons for Brazil in the successes and failures.

13. First, when the amount of information required – as determined by the number of programs and the information sets – becomes too great, the programs are ignored. This is what happened in Australia’s first reform (1984) and in the State of Hawaii. Australia’s revised approach of aggregating activities into a small number of portfolios, worked as a device for giving managers freedom to manage and making them accountable. Brazil’s PPA appears to stand near to the Hawaiian experience.

14. Second, in the review of these cases, it is axiomatic for the authors that, in the real world, it is politics and practicality that have meant that budgets continue to be appropriated by organizations: “it is exceedingly difficult to subordinate the organization structure, even when the government has a program budget, because organizations actually spend the money and are responsible for results.” (page 27). If programs are to be used to improve policies (and reduce redundancies), the example of the Department of defense indicates that this must be done separately from implementing the policies, which is the province of organizations.

15. Perhaps the most important inference for Brazil is that the PPA has made the mistake of trying to reach two objectives – policy making and management incentives— with one instrument. The Department of Defense succeeded in the one, Australia in the other. To compound things, Brazil was also trying to achieve a third objective, integration of plan and budget. Brazil has chosen a more ambitious path than any other country in the world, with virtually no exception.

APPENDIX 2: THE EXPERIENCE OF PROGRAM AND POLICY EVALUATION IN SOME OTHER COUNTRIES

16. The systematic evaluation of government policies and programs is part of a broader toolkit for public-sector management, and it is usually considered to sit within the paradigm of “performance-oriented” management. Indeed, donors persistently recommend evaluation, within this paradigm, and very often build it into the projects they finance.

17. Yet the empirical basis for understanding when and how evaluation is likely to be effective seems to be thin.²⁴ Gradually, over the last two to three decades, evaluation has become normal practice in the governments of most OECD countries, but evidence about the contribution of evaluation is at best limited and unsystematic. Only one OECD country, Australia, has been generally acknowledged to have clearly benefited from systematic evaluation – yet it has now partially abandoned this system. Evaluation is less practiced in developing countries, but several Latin American countries have set up evaluation systems. Chile, in particular has established an effective evaluation system that can be compared to Australia’s.

18. This note seeks to summarize what we know about evaluation and to put evaluation into a broader public-management reform context. These are the principle questions to be asked:

- What is the benefit of formal evaluation? How does it fit into a broader picture of how governments learn and make policy?
- Where does evaluation fit within a public-management-reform sequence?
- What are the circumstances under which evaluation is most effective?
- How are evaluation activities best organized?

19. The available evidence is too thin for strong conclusions to be made about the do’s and don’ts of evaluation. Rather the intention, is to provide a framework to help raise questions about the role of evaluation in the continuing process of public-sector modernization, especially in developing countries.

20. The note will proceed by: providing some definitions and raising a few conceptual issues about where evaluation fits into public-sector management and reform; briefly reviewing the evaluation experience of some leading OECD countries; and, finally, assessing the important issues that arise.

²⁴ The empirical literature describing and analyzing real-world cases of public-sector evaluation is limited. It is largely written by experts from the evaluation community, a community that is very active in the OECD countries. The mainstream performance-oriented-management literature is (which is much more copious) does not dwell much on the details of evaluation. The main relevant point it makes is to emphasize the role of evaluation in the budgetary decision-making cycle.

A. Evaluation and Public Sector Management

21. *Definitions.* An evaluation is a formal analytical exercise to judge the results of a program or policy. Has it been, or is expected to be, efficient, effective, and sustainable? (See a more formal definition in Table 2.1.) Formal evaluation is usually ex-post.²⁵ Good evaluation uses, as much as possible, hard data and analytical methods that posit a “policy theory” (or “program logic”), a theory of causality that explains how inputs get turned into outputs and affect outcomes.

Table 2.1. Monitoring and Evaluation: Definitions

	Definitions	Comments
Monitoring	A continuous process of collecting and analyzing information to compare how well an activity, project, program, or policy is being implemented against expected results.	Monitoring can be undertaken at various levels of sophistication. At the simplest level, actual outcomes can be compared to expected outcomes. But “results-based monitoring” would seek to relate outputs and outcomes, in addition, to inputs and processes. A monitoring system is often formalized and automated within a Management Information Systems (MIS). ²⁶
Evaluation	An assessment of a planned, ongoing, or completed intervention to determine its relevance, efficiency, effectiveness, impact and sustainability. The intent is to incorporate lessons learned into the decision-making process.	Evaluation exercises are typically occasional and ex-post. Ideally, they use analytical methods that involve hard data and explicit models of causation. ²⁷ There does not seem to be agreement on when “softer” approaches to understanding the cause and effect behind policy implementation – such as more general sector studies or the more casual empiricism of journalists, politicians and bureaucrats – start or stop counting as “evaluation”.

22. Evaluation is often associated with monitoring; indeed, monitoring and evaluation (M&E) are usually considered to be closely linked information-seeking activities. Monitoring is a regular process of checking information to see whether an activity, program, or policy is being implemented according to expectations.

23. *The uses of monitoring and evaluation.* The systematic acquisition of information (learning) through M&E serves a variety of management uses, notably ensuring that agents are performing as they are supposed to, seeking ways to improve the efficiency of resource-use, and seeking a better distribution of resources between different activities. In practice, evaluation tends to relate more to alternative uses of public resources in the longer term and monitoring more to management control in the shorter term. But it is

²⁵ Ex-ante evaluation goes under other names, typically “appraisal”.

²⁶ A Management Information System (MIS) is a system, usually computerized, for collecting, processing, and distributing up-to-date information, thereby facilitating an organization’s systematic access to program and participant information. The types of information typically included in a MIS include inputs, processes, and outputs (and sometimes outcomes), classified by program. These systems provide information in a form that managers can use to make decisions at all levels of the organization: strategic, tactical, and operational. A typical MIS report is a periodic (scheduled) report for a principal. Many MISs can be adapted to meet evaluation requirements.

²⁷ See World Bank, OED (2002).

difficult to say where monitoring stops and evaluation begins. Running a data base of education statistics is obviously a monitoring activity. A study of why a family-subsidy program did not lead to better school attendance is obviously an evaluation activity. In between are a host of activities – descriptive studies, investigative newspaper articles, management reports, plain *ad hoc* interpretation of number, and so on – which are more than monitoring and less than formal evaluation.²⁸

24. *Evaluation and public-sector reform.* Monitoring and evaluation, especially inasmuch as they are linked activities, are most frequently represented as part of a modern public-expenditure-management (PEM) system. In a conventional PEM system, the decision cycle is typically limited to one year, the planning and implementing unit is the agency, the emphasis is on the legal control of expenditures, budget decisions are typically incremental, and managing for results and learning have a low priority.

25. In the idealized form of a modern system, policy-making and -coordination are more rational, financial planning extends beyond one year, information based on programs becomes a tool for performance management, and learning about past performance is institutionalized, all within a virtuous circle of public expenditure.²⁹

- *Policy making:* policies are prioritized (implicitly at least in cost-benefit terms) and their aggregate costs respect an overall resource constraint.
- *Medium-term framework* (planning): spending is planned ahead so that programs can be rationally and economically implemented (notably through coordinating current and capital expenditure) while respecting expected future resource constraints.
- *Program accounting* (information): inputs are fully costed, outputs are specified as concretely as possible, and cost-accounting enables the calculation of efficiency (output per input).
- A *feedback* mechanism (an M&E system): performance is monitored to ensure that it is in line with expectations and results are monitored in order to improve future policies and implementation.

Thus an M&E system is generally thought of as part of a performance budgeting system, itself part of the toolkit of performance-oriented management.

26. Notwithstanding this conceptual role of M&E, it is not clear that monitoring and evaluation *need* to have an umbilical relationship. Nor is it clear that evaluation needs to be embedded in a *system*.³⁰

- On the one hand, M&E are conventionally put in the same “box” because they both pertain to the systematic use of information to improve performance, they overlap in

²⁸ Learning can be a very incremental process: discovering new facts leads to new puzzles which need solving either in the light of new facts (more monitoring) or through formal analysis (evaluation).

²⁹ See Figure 3.1 of World Bank (1998) PEM Handbook for a diagrammatic presentation of the public-expenditure-management cycle.

³⁰ An evaluation system is a government-wide or agency-wide system of uniform rules and practices on evaluation that enable a basis of comparison of results and recommendations and control of quality across programs (agency-wide) and organizational units (government-wide).

functions, and it is difficult to draw a dividing line between them. On the other hand, each belongs to other “boxes”. First, the salient use of monitoring is to provide management information (often in real time). Second, evaluation is also linked – equally strongly, surely – to a different “box”: the broader domain of policy-making (see below).

- Obviously, if cost is no object and rationality is unbounded, systematizing the evaluation function will encourage evaluation planning, routine evaluation, the comparison of evaluation results, and the maintenance of evaluation quality. But systems can exist at various levels: at its most sophisticated, an evaluation system would be integrated with a performance-budgeting information system where inputs are fully costed and results fully specified on a program basis. (Australia, then Chile, provide the nearest examples of this.)

Evaluation *can* be carried as a non-routine activity (as it is done in donor-financed projects, for example). And it *can* be carried out independently of a sophisticated budgeting system.

27. *Evaluation and policy making.* Evaluation is just one part of a complex web of policy-making tools, rules and institutions (Box 2.1). The policy-making function is not well understood (or else it is taken for granted) in the public-management literature, but there is no evidence to suggest that formal evaluation plays anything but a modest role overall. Should evaluation play a larger role? It is logical to suppose that an analytical approach to policy-making will produce outcomes that are more in the public interest than political or intuitive decision-making or the views that emerge from the interactions of individuals and organizations within policy networks?

28. The growth of formal evaluation in the OECD countries supports this view. But there will be limits to the role evaluation can play for several reasons.

- First, there is a technical limit to what even the best evaluation can do. Formal analytical techniques are likely to prove better in asking questions related to effectiveness and efficiency – how or how well is the activity contributing to meeting its objectives? – than at asking questions related to costs – how much does an activity cost compared to its benefits?. This is because cost-benefit analysis is particularly difficult to apply across the broad range of activities of the public sector (Box 2.2). Cost-benefit analysis can sometimes be the basis of individual project decisions, but it is not feasible for the kinds of choices between policies that have to be made at the center of government.
- The second limitation is politics. Even in the best-run, most enlightened governments, decisions that cement politicians’ constituencies will continue to be central to politicians’ policy preferences.
- Finally, there is a particular argument regarding evaluation *systems*, not evaluation as a tool. This is part of a broader argument about the complex systems involved in running a modern performance-oriented public-management system. Not only do such systems involve high transactions costs in collecting, processing, and reporting on performance and accountability-related information. They also run the risk of

providing too much information for their operatives to digest and use (the so-called problem of bounded rationality). If policy is mostly made through intuitive and political decisions and through the paradigms that emerge from interactions within policy networks, this is perhaps because human and organizational capacities are mostly insufficient to do any better. (But this should not provide an argument against reasonable efforts to make rationality less bounded.)

Box 2.1: Evaluation and Policy-Making

Formal evaluation and the systems that support it occupy just one part of a broader canvas of policy-making tools, rules, and institutions.

The tools of policy-making. Formal evaluation (i.e. analytical study of a particular activity) is just one of many inputs into informed policy making. Other inputs are: dogma and theory (belief sets are important when empirical information is not available or there is no agreement on causal models); learning from foreign experience; experiment (market testing, for instance); and casual observation or information from users

The rules and institutions of policy-making. A requirement for formal evaluation (i.e. the routinization of evaluation) and the formal linking of evaluation to decision making (for instance, a rule that new budget proposals must be accompanied by evaluative materials) is just one possible part of a more general infrastructure of rules and institutions to support policy making. Other policy-making rules might include requirements for pre-publication and public discussion of new policies, as well as regulatory impact assessments. The institutions of policy-making are complex and varied. There are formal mechanisms: the budget-decision cycle incorporating formal learning and review (with evaluation systems possibly part of this); occasional program or spending reviews, regular deliberative and decision-making bodies (notably Cabinets and committees); policy think-tanks and special-purpose or *ad hoc* commissions and enquiries. But it is no doubt informal policy networks – communities of stakeholders with expertise, inside knowledge and power – that are the most important in affecting the course of policy change. Such networks are comprised of political interests, experts employed in the sector, consultants, academics, and so on. Policy networks allow a discourse – a story line – or competing discourses, on what needs doing and how.

Political scientists have written much about policy-making, but their findings have not entered far into the sphere of the public-sector-management community, especially as regards developing countries. Thus, in discussions of public-sector organization, policy-making remains something of a black box. Evaluation is part of the puzzle, but perhaps it is only a modest one.

B. The Experience of Evaluation in Selected Countries

29. The easily accessible literature on individual country experiences with evaluation systems is thin and often not up-to-date. The following paragraphs comment on a small (and somewhat ad hoc) selection of national experiences where evaluation systems exist, with some degree of effectiveness at least. The countries covered are Australia, Canada, the UK, and Chile. The major differences between the national systems concern the extent to which the center of government provided guidance and/or authority, the degree of discretion left to the agencies, and the type of impact that evaluation has had (typically, whether it has had more effect on policy prioritization – “Level 2” – or on operational effectiveness – “Level 3”). Table 2.2 seeks to summarize the differences for these four countries.

Box 2.2: Policy-Making, Cost-Benefit Analysis, and Evaluation

Formal (ex-post) evaluation, as well as (ex-ante) appraisal, involve many different methodologies. But any one of these methodologies responds to one of two questions:

- how much does the cost of an activity compared to its benefits? or
- how/ how well is the activity contributing to meeting its objectives?

The first question is addressed, ideally, by cost-benefit analysis, the second by theory-based evaluation.

Cost-benefit analysis compares the costs and benefits to society of an activity in monetary terms, using shadow prices to account for externalities. In practice, cost-benefit analysis for public-sector activities is much more difficult than for the production of marketed goods and services, which have observable prices and, unlike public goods, limited externalities. Cost-benefit analysis may be useful in deciding on large public projects (such as a dam or a military airplane) or in deciding on alternative options to reach the same objective (the cost of a life saved through seat-belt regulation versus crash barriers). But it is not very useful, in practice, in comparing a large amount of alternative public-expenditure options.

Theory-based evaluation constructs (and sometimes tests) an explanation of a causal relationships: the process by which inputs get turned into outputs and how outputs are related causally to outcomes.

Evaluation in reality covers a good deal more types of analysis than theory-based evaluation. There is a whole range of formal-to-informal analytical methods which throw light on issues of efficiency, effectiveness, and cost and benefits – Public Expenditure Tracking Surveys, for instance.³¹

Because they answer different questions, cost-benefit analysis and theory-based evaluation are likely to play different roles in policy-making. Cost-benefit analysis is about the social utility (profitability) of one use of (public) money compared to other uses. It does not tell us whether there is/was a more efficient way to reach given objectives. Theory-based evaluation is about the choice of “technology” to reach an objective compared to other choices to reach the same objective. It does not tell us whether the activity is/was a justified use of public money.

It follows that theory-based evaluation (indeed, any form of evaluation which does not monetize costs and benefits) is unlikely to be very useful in comparing different policies – in deciding, say, whether an extra dollar spent on education expenditure is better than using that dollar on defense. But within a sector where there is some substitutability between activities – pre-school learning and primary education, for example – evaluation may help policy-makers make spending choices.

Unfortunately, cost-benefit analysis cannot do this for lack of data. So the kind of policy-making that has to be done at the center of government (“level 2”) lacks for analytical techniques and has to rely more on judgments based on theory, on political beliefs, or on experience culled from elsewhere.

Australia³²

30. *The evaluation system.* Australia experienced two distinct periods of performance-oriented management from the late 1980s (see Appendix 2). Systematic evaluation flourished when the federal administration’s emphasis was on *improving government* through more coherent policies (1987-97) and weakened, from 1997, when the emphasis shifted to *less government* (by reducing the size of government, simplifying formal requirements, contracting out, and so on).

³¹ See World Bank, OED (2002) for an introduction to principal tools, methods, and approaches to M&E. The causality model bears some relationship to the economist’s production function: a mathematical relationship between the quantity of output of a good and the quantity of inputs required to make it, based on an assumption about the technological relationship of input to output. The Logical Framework (LogFrame) is a stripped-down approach to theory-based evaluation.

³² The account of Australia relies largely on MacKay (2003). See also comments on the evaluation system in Schick (2001).

Table 2.2: A Comparison of National Evaluation Systems

<i>Public-management context</i>	<i>Role of center of government</i>	<i>Role of agencies</i>	<i>Outcomes</i>
Australia, 1987-97			
Performance-management system emphasizing policy coherence	Strong central control of evaluation system by strong finance ministry, backed by strong incentives.	Operation of evaluation systems under central guidance.	Effective use of evaluation findings in budget decisions.
Canada since 1977			
Sharpened focus on performance and prioritization in recent years.	Procedural rules set at the center.	Departments run their own evaluation systems.	Evaluations tend to focus on issues of operational effectiveness (rather than broader issues)
UK since 1977			
Performance-management system emphasizing value for money, but with more recent emphasis on prioritization.	Methodological advice from center. Weak links to other performance instruments (till recently)	Substantial discretion in nature of evaluation systems.	Uneven results. Generally a weak role for evaluation.
Chile since 1997			
Performance management system aimed at efficiency in resource allocation.	Strong central control of evaluation system by strong finance ministry.	Evaluations carried out by center. Agencies are relatively passive.	Evaluations contribute to budget decisions and operational effectiveness.

31. The government adopted an ambitious formal evaluation strategy in 1988. Its objectives were to support Cabinet decision making (policy prioritization), help departments improve the performance of their programs, and provide formal evidence of accountability. All programs were to be evaluated by the departments on a three-to-five-year cycle. Most of the evaluations were published and their results were reflected in the budget documents. The system was centrally driven by a very strong Department of Finance and Administration (DoFA) which oversaw methodology, design, and publication of results and effectively promoted departmental evaluation through a system of “carrots” and “sticks”. The carrots that encouraged the departments to take evaluation seriously were the offer of advisory services from the DoFA and the opportunity for the Departments to use evaluation results as part of their arguments for justifying budget resources. The sticks were the DoFA’s influence over budget allocation and its influence over the reputations of Departments. Australia’s system is thought to have succeeded because it gave affected departments a big stake in designing and using evaluations (Schick, 2001).

32. From 1997, a new administration eliminated most formal evaluation requirements, as well as and technical support from DoFA. Instead, performance measurement (outcomes and outputs) and incentives associated with these have become the basis for achieving the aims of performance-oriented management. (However, some programs still require a performance review to justify their continuation, while some departments carry out good evaluations for internal purposes.) With the decline in

evaluation, the quality of performance information has reportedly become a problem. This period has also been characterized by a reduced oversight role for DoFA.

33. *The uses of evaluation, 1987-97.* In the earlier period, the evaluation system was part of a well-articulated performance-management system emphasizing policy prioritization. For instance, from 1987 all new spending proposals had to specify their objectives, propose how their performance would be measured, and establish a plan for ex-post evaluation. Many of the subsequent evaluation studies identified the need for better performance measurement. This initiated a series of departmental reviews, from 1995, which led to better focus on performance information. There is also clear evidence that evaluations were used intensively in the budget process, with respect to trimming existing programs and justifying new ones. (DoFA published an annual report estimating the percentage of that year's decisions influenced by evaluative findings.) MacKay (2003) quotes the observation of the Auditor-General that "In my view, the success of evaluation at the federal level of government was largely due to its full integration into the budget processes. Where there was a resource commitment, some form of evaluation was necessary to provide justification for virtually all budget bids."

34. The Australian case suggests that the well-articulated combination of a strong central authority (making rules and applying pressure), decentralized implementation, agencies' freedom to manage, and fiscal room for policy-based (rather than zero-based) budgeting were the necessary conditions of effective evaluation. But notwithstanding the apparent success of evaluation in 1987-97, the new 1997 government was not convinced of its utility.

35. MacKay (2003) has drawn some generalizable lessons from Australia's evaluation experience which are summarized in Box 2.3.

Box 2.3: Australia's Evaluation System – Implications for Developing Performance Systems

- A balance of "carrots, sticks and sermons" works best. The 1987-97 evaluation strategy worked because it was centrally driven: a strong DoFA used a mix of carrots and sticks, through, ultimately, carrots and persuasion proved more important than sticks.
- It took some time for the DoFA to convert from an interest in controlling spending to a pursuit of value-for-money spending.
- The policy-making process, organized around the annual budget cycle, is a powerful vehicle for concentrating on performance.
- A simple advocacy of evaluation is not enough: reform must be spearheaded by a ministry champion. But it also helps to have other agencies – notably, in Australia's case, the Australian National Audit Office – support this.
- Persistence and commitment are needed, as it takes years to develop and fine-tune capacity. Reforms should be institutionalized – and they are easy to undo than to do, as the post-1997 experience shows.
- Reform must be demand-driven.
- Reform is not just a technical issue; it is also cultural.
- An evaluation system should not be over-specified and over-engineered.
- Australia's self-evaluations avoided subjectivity by the use of specialist units, oversight committees on more important evaluations.

Source: MacKay (2003)

Canada³³

36. In the 1970s, Canada's federal government organized an ambitious evaluation system around the Comptroller General. The effort bore little fruit. The system is thought to have failed because it centralized evaluation, thereby dampening cooperation from the spending departments which feared they would be adversely affected by the findings (Schick, 2001).

37. Since this time, an alternative system of internal evaluation has become firmly established. A new evaluation policy was set in 1977, revised in 1994, then updated in 2001.³⁴ The policy makes evaluation a routinized management tool, within a broader framework of performance-oriented management. The process is decentralized. The Treasury Board of Canada (TBC – a group of ministers, supported by a Secretariat, that acts as manager of government) advises on best practices, sets standards, monitors the evaluation capacity in the departments, and uses the products of evaluation to inform decision-making at the center. The departments and agencies (numbering over 70) are responsible for organizing their own evaluation systems, undertaking evaluations, using their results, and sharing their findings with the TBC. The 2001 Evaluation Policy is not detailed in terms of procedures, but it requires that evaluations take place within an evaluation framework that defines the program's *raison-d'être*, how it should work (its program logic), how performance will be measured, and how evaluation will be undertaken. The government has sharpened its focus on performance since the later 1990s: public-service managers are accountable for results, responsible for reporting accurate results; and evaluation. The Report on Plans and Priorities tabled prior to each fiscal year, provides details on business lines for next three years; the Departmental Performance Report, tabled after the end of the year, provides information on results and costs.

38. The Canadian evaluation system is, thus, one where different departments are free to apply the policy with more or less energy, depending on their sense of the usefulness of evaluation. Over 25 years of the system has made good, but sporadic progress. After some years of cutbacks, the evaluation function is currently being strengthened. Mayne (1994) used survey data to establish that evaluations do, indeed, often have consequences for programs, but most of the resulting changes are of the "single-loop-learning variety (doing the same thing better) rather than "double-loop-learning" (i.e. more fundamental changes in program design). In a similar vein, Divorski (1998) talks of evaluations having their impacts in more limited areas of operational effectiveness, rather than broader strategic areas.³⁵ There is room to improve results measurement capacity, as well as the link of evaluation to policy making.

³³ See Mayne (1994) and McDonald (2002).

³⁴ See http://www.tbs-sct.gc.ca/pubs_pol/dcgpubs/tbm_161/ep-pe_e.asp.

³⁵ Divorski (1998, page 71) also reports that, in an investigation looking at a number of countries, "we found that when the responsibility for decisionmaking about evaluations is located close to program

The United Kingdom

39. A systematized attempt to introduce evaluation in the 1970s foundered, at a time when overall expenditure levels were a greater source of concern than value-for-money spending. However, this initiative implanted the beginnings of an evaluation culture. There was a growing emphasis on efficiency in the 1980s (see Appendix 2). The government took an important step to promote policy evaluation as a regular and organized (but decentralized) activity with a 1985 decision that all Cabinet proposals with value-for-money implications would have indicate what would be achieved, by when, and at what cost, and how this would all be measured. The role of the center of government is largely limited to methodological advice to ensure consistency in intellectual frameworks.³⁶ In this system, the individual departments play a leading role. They tend to be organized for evaluation in different ways.³⁷

40. One study, as of the late 1980s, reports that that evaluation practices were fragmented, reports were seldom published, and parliamentary scrutiny was not well advanced (Jenkins and Gray, 1990). This relatively weak role for evaluation may have begun to change from late 1990s, as the budget came to be promoted as the leading instrument for performance management (see Appendix 2). The evaluation of policy effectiveness became a specific criterion for fund allocation. The 2002 Spending Review required that the most ineffective or unimportant five-percent of a department's programs be wound down over 3 years.

Chile³⁸

41. Chile's public administration is one of the best in the non-OECD world. It is effective in delivering public services, its public servants are competent and professional, and its operations are predictable. Since Chile's return to democracy in 1990, successive governments have emphasized fiscal discipline, social equity, and, increasingly, good management. Chile has a highly centralized and disciplined system of public financial management. The period since 1990 has been one of remarkable fiscal discipline and equilibrium. It was in this context that the government began to incorporate performance-oriented instruments in public management and financial decision making, including performance indicators, management improvement programs (which are now being applied to all agencies), and annual performance reports to Congress. Chile's performance-oriented instruments are overseen by a powerful budget office.

management, the evaluations tend to focus on operational effectiveness issues and tend to limit the examination of impact issues.”

³⁶ The only central guideline on evaluation is contained in the Treasury's Green Book, first published in 1977 and updated in 1991 (see Treasury, H.M. (1997).

³⁷ The Department of Trade and Industry is one of the leading departments in evaluation practices. It has organized its own evaluation system, which starts with an evaluation planning exercise.

³⁸ This section is based on World Bank (2006).

42. As one of these instruments, Chile began to develop systematic evaluation of government programs from 1997. The objectives of the system are fourfold: to help the budget office formulate the budget, to help Congress discuss it, to improve program management, and to improve public accountability. It appears that Chile has established a sophisticated and effective system which produces good results and ensures that evaluation recommendations are acted on.

43. The evaluation system is characterized by several salient features. First, it is highly centralized, disciplined, and routinized and offers little discretion to evaluators. Second, it relies on the expertise of outside evaluators secured by transparent contracting processes. Third, for the bulk of the evaluations undertaken, methodologies are simple (using a Logical Framework that emphasizes outcomes), they are not costly, and reports are produced quickly. Fourth, evaluation recommendations systematically form the basis for agreements between the budget office and the agencies evaluated on corrective management actions. This happens within an integrated, routinized, and centrally-controlled set of activities to report on and improve performance.

44. This system has since 1997 produced a quantity of evaluations that now covers a good proportion of central-government programs. It is claimed that evaluations are having several beneficial impacts. First, the budget office claims that evaluation results, as one of several non-binding input, improve budget decisions. (There is so far little evidence that the Ministries and Congress make similar use of evaluation results for budget or policy decisions.) Second, agencies are required by the budget office to act on the agreed corrective management actions. Third, the process of being evaluated leads some (not all) programs or agencies to adopt a more strategic view of management.

45. Chile's evaluation system is effective (at the very least in the sense that it routinely produces evaluations that are acted on) because of particular elements of system design and favorable external conditions. In terms of system design: evaluations are produced according to a well-designed production routine (in effect, Chile has created an evaluation "factory"): the use of external evaluators creates credibility; there is a management system for ensuring that results are used.

46. Several external conditions underpin the evaluation system. First, Chile has an extraordinarily powerful budget office. Second, Chile's good financial management is able to reconcile fiscal-equilibrium and quality-of-results objectives. Third, public servants are highly professionalized, and they are beginning to be imbued with elements of a performance-based culture.³⁹ Finally, there are particular political conditions in Chile (underpinned by fears of returning to some of the problems Chile faced in the decades before 1990) that favor the application of neutral expertise to management and policy problems, i.e. that "technify" many political decisions. This combination of enabling factors distinguishes Chile from virtually all other countries.

³⁹ Some claims are made (and these should be subject to confirmation) that Chile's incipient results-based-management practices are largely driven by the popular expectations and demands that accompanied Chile's return to democracy.

C. The lessons of country experience: main points

1. The evidence that evaluation, let alone systematic evaluation, is a key component of performance management is hard to come by. The practice of evaluation has grown substantially in the governments of the advanced countries in the last two to four decades. But the sparse literature (largely the product of the evaluation community) has few unconditional stories about the success of evaluation in changing programs or policies on a significant scale.
2. By the same token, there is no evidence that evaluation has become a key – or priority – element in the broader canvas of policy-making. A better understanding of this broader canvas may well be a pre-requisite for a better understanding of what evaluation can do.
3. But since there is such a large process of change going on in public management practices in the more advanced countries, these observations do not necessarily indicate what role evaluation could, or ought to, play.
4. Evaluation answers some policy questions better than others. It works best when agencies want to know how well their programs are doing (“single-loop learning”).⁴⁰ It is not a proven instrument for cross-agency comparisons, or for big-picture policy decisions (“double-loop learning”) and spending-review-type decisions that involve alternative use of funds.⁴¹
5. Many experts believe that evaluation is best carried out by the agencies themselves, rather than the center of government (because of the incentives and processes involved in learning). Australia’s success (1987-96) came through a judicious mix of sticks and carrots (centralized rules, decentralized implementation). (External comptrollers often do good evaluation, but it is not clear how well the results are used by the agencies.) Chile’s experience runs counter to the notion that decentralized evaluation works best, but Chile’s central office (the budget office) is extraordinarily powerful.
6. Should countries invest significantly in evaluation capacity, especially in a government-wide evaluation (or M&E) *system*? It is perhaps premature for most countries – including most advanced countries – to assume the need for a large investment in a government-wide evaluation *system*. The international comparisons suggest that an evaluation *system* makes most sense when:
 - a. there is a performance culture;
 - b. the central reform emphasis is on policy coherence (but this is based on a single observation: Australia 1987-96);

⁴⁰ An OECD symposium on knowledge management in central government (OECD, 2003) concluded that evaluations are effective in improving management of public programs, but do not produce major policy change. Apparently, such major changes require (public and politically-brokered) value changes.

⁴¹ Furubo (1994) concludes that, for Sweden, evaluations were central to technical adjustments to existing programs, but rarely played a role in changing policies.

- c. major elements of a performance-based budgeting system (including some flexibility in resource use) are in place.

We can usefully make this same observation in the context of Schick's (2003) sequence of performance measures : first, you certainly need to focus on performance and you probably need, Schick suggests, organizational changes that will encourage and underpin performance measurement. Perhaps it is more important to first think of evaluation as a *culture* before we think of it as a *system*.

- 7. Developing countries are far from being at this point in respect of policy-making and performance orientation. (Chile is an important exception.) The developing countries do not appear to be driven towards performance management by the same crises of performance and legitimacy that have driven reform in the advanced countries. The challenge for developing countries is more akin to that faced by today's advanced countries in the Nineteenth Century. And prudent sequencing suggests caution (Schick, 2003).
- 8. National political arrangements may affect the nature of demand for evaluation and the way systems are implemented. External evaluation (by auditors answerable to parliament) may be more important in presidential systems, internal evaluation in Westminster systems. Evaluation may be technically more challenging in federal governments inasmuch as federal outputs are distant from sub-national outcomes.

APPENDIX 3: THE QUALITY OF PPA EVALUATIONS

The Quality of PPA Evaluations

47. PPA evaluations have been regarded as ineffective by program managers and M&E policy community in Brazil. To verify this widespread opinion, an analysis was conducted to compare PPA evaluations and TCU's operational evaluations (*Avaliações de Natureza Operacional*). From 1999 to 2006, TCU has undertaken 50 evaluations of selected federal programs. TCU evaluations are considered of good quality, with additional merit of influencing the whole government towards commissioning external evaluations to improve program management.

48. A comparative analysis was undertaken for nine PPA programs, for which TCU has already conducted its operational evaluations.⁴² For each program, five years of self-evaluations (2000-2004) were compared with the methodology, data and recommendations made by the corresponding TCU evaluation. TCU evaluations involved interviews of federal managers to get a better understanding of the program's logic. In addition, they also include fieldwork to interview beneficiaries and municipal managers and collect output information. Hence, the methodology applied by TCU evaluations is more complex than PPA.

49. Self-evaluations are intended to help managers themselves identify and solve their problems. According to the managers we interviewed, however, the PPA evaluations have not been effective in this function. Every year program managers are requested to self-evaluate, but there is no feedback from the Ministry of Planning regarding solutions.

50. The format and quality of evaluations has varied very little along those years. They are divided in four parts: Results, Design, Implementation and Recommendations. The evaluations are followed by three annexes containing: (i) physical/financial execution of actions, (ii) PPA indicators and (iii) costs. In the report's initial section, managers provide programs' aggregate outcomes/activities (e.g., coverage, financial execution, number of facilities built, number of events). The data is general for the country and lacks program-specific details such as geographic distribution of the services provided. Managers usually do not analyze the potential impact of these results for beneficiaries' lives.

51. For the great majority of programs the program design is considered adequate. Overall, criticisms tend to be directed toward other ministries or municipal managers who implement the programs in Brazil's decentralized setting. Managers also criticize some details, such as the program's stated objective, targets and action's names. Hardly ever managers analyze the program logic, even when they are not executing the budget satisfactorily.

⁴² Selected programs are: Desenvolvimento do Turismo no Nordeste, Saúde da Família, Atenção à Pessoa Portadora de Deficiência, Saneamento Básico, Energia das Pequenas Comunidades, Programa de Erradicação do Trabalho Infantil, Educação de Jovens e Adultos, Novo Mundo Rural, Morar Melhor.

52. Recommendations are suggestions that managers do for themselves. Usually, these suggestions are not self-critical, but are general statements of what should be done, without including possible concrete solutions. There is no incentive for managers to reflect on the causes of their problems since there is little chance they will be addressed. The most recurrent issues are: discontinuity of financial flow due to *contingenciamento*, problems of coordination with other implementing ministries, lack of infra-structure and personnel capacity.

53. In general, PPA indicators are not realistic and measurable. Managers do not provide relevant quantitative/qualitative indicators, mostly because they have poor monitoring systems and there is little effort to seek information directly from municipalities, except by PETI manager that send out questionnaires to municipal managers. Budget information provides general figures for execution, but it is detached from reality because it does not capture common issues of Brazilian budgetary process, like budget unpredictability (chronogram of disbursements) and arrears executions. Some managers commission external evaluations from research institutes, bringing selected results to the self-evaluation (e.g. opinion survey of *Programa Atenção à Pessoa com Deficiência-2001*). This is a good practice and should be encouraged by SPI.

54. Overall, PPA evaluations are characterized by lower quality and coverage than TCU operational evaluations. This is mostly due to methodological differences. TCU evaluations use several methodological tools to obtain and judge information about programs (document/data analysis, direct observation, focus groups, mail questionnaires and interviews with beneficiaries). Program managers could make an effort to use data that is available from IBGE researches like PNAD to improve evaluation quality.

55. Nevertheless, the majority of PPA evaluations are able to identify some shortcomings that may later be uncovered by TCU. However, they lack specificity, analysis and, above all, ideas on how to solve problematic issues. It is often possible to find recurrent inadequacies, year after year. In contrast, TCU recommendations are specific and well reasoned. However, they usually do not present solutions to problems either, recommendations (e.g., “*seek more integration with other ministries*”) usually do not carry along “how to” solutions to tackle the pinpointed problems. Managers are obliged to comply with TCU’s recommendations and programs are monitored for two years after the operational evaluations.

56. When comparing TCU and PPA evaluations, it is possible to note consistency in the identification of problems for six programs (*PETI, Saúde da Família, Saneamento Básico, Novo Mundo Rural, Morar Melhor, Atenção à Pessoa com Deficiência*). Three programs demonstrated much lower capacity (*Energia nas Pequenas Comunidades, Turismo no Nordeste, Educação de Jovens e Adultos*.) Even though PPA evaluations were able to pick up evident problems, TCU was able to arrive at more nuanced and relevant findings which were not captured by self-evaluations (e.g., learning difficulties and high repetition rates of EJA courses, purchase of electric generators for communities that were already served by electric network services, significant part of *Morar Melhor* beneficiaries sell their houses, INCRA’s serious organizational deficiencies). Moreover, TCU was able to articulate these problems in a more effective way.

57. Another import issue is the capacity that managers have to solve problems from one year to the next. Again, for some programs the evolution of improvements across years is clear from the self-evaluations (*PETI, Saúde da Família*). For others, managers bring up the same problems year after year (*Novo Mundo Rural, Turismo no Nordeste*). Finally, for others changes occurred after the TCU evaluation or the program was restructured after the new government took office in 2003 (*Energia nas Pequenas Comunidades, Educação de Jovens e Adultos, Morar Melhor*). No manager has ever attributed changes to the self-evaluation report or Ministry of Planning support, at least in the sample of the programs reviewed for this analysis. On the contrary, the most common issue that is object of complains remains: discontinuity of financial flow due to *contingenciamento*.

PETI – Programa de Erradicação do Trabalho Infantil (2000-2003) (2004-2007)

58. *Programa de Erradicação do Trabalho Infantil* (PETI) aims at eradicating child labor (children under 16 years old) through monthly cash transfers to the family and keeping children in after-school activities. From the analysis of five PETI self-evaluations (2000-2004), it is possible to understand the logic, large expansion, and the importance of civil society organization for the program. However, the quality of evaluations has varied along the years. In 2000, the information was very superficial. In that year, the manager contracted an external evaluation, but the results were not incorporated in the text. Since then, the quality of information has improved, due in large part to the manager's effort to send questionnaires to municipalities and get coverage data from PNAD (the results of which were used in the 2001 and 2002 evaluations). Nevertheless, challenges still remained, such as the ability to check information *in loco* and the absence of a reliable diagnosis of child labor incidence.

59. Overall, evaluations illustrate PETI's positive results but also its shortcomings. Problems are pinpointed (delays in transfers due to *inadimplencia* of states with INSS, weak oversight due to shortage of labor inspectors, lack of infra-structure and capacity for *Jornada Ampliada*, etc). Along the years it is possible to see how some of the shortcomings were solved (direct transfer to families via electronic card, public tests to hire labor inspectors) and how others were not. Recurrent problems, highlighted by evaluations are: lack of human resources, capacity, infrastructure, social control and monitoring in municipalities. In general, the managers do not provide ideas on how to solve their recurrent problems.

60. Hard information presented by action on budget/execution (annex) is fairly complete and coincides with the data presented in the text. However, for each of the evaluations over the five year period, program indicators (child labor rate, rural and urban) were never provided in the annex, even though they were listed in the text in both 2003 and 2004. Cost information is more irregular. On occasion, the cost is given, but the units/products are unclear (e.g. *concessão de bolsa criança cidadã*- R\$ 339,00, where

the unit is *criança atendida*).⁴³ For other actions, products are indicated as “não aplicavel” or “sem execução”.

61. PETI evaluations could provide an interesting case to improve government-wide management on how to solve duplications. The merger of PETI’s cash transfer payment system with Bolsa Família’s, facilitated the migration of families to *Cadastro Unico*. This was a complex operation that could have been better explained in the 2004 evaluation (i.e., if clearly explained, the PPA evaluation could have served for learning purposes).

62. The recommendations for all the years were very consistent with TCU findings. Overall, TCU found very positive impacts of the program, using more complete methodology with almost 100 interviews with managers and families in six states (BA, MA, MS, PR, PE, SE) and mail questionnaires sent to managers and committees in 967 municipalities.

Desenvolvimento do Turismo no Nordeste PRODETUR I e II (2000-2003)/Turismo no Brasil: uma viagem para todos (2004-2007)

63. PRODETUR aims at developing tourism infra-structure (e.g., roads, airports and sanitation) in the Northeast region. The initial program evaluations (2000 and 2001) provided very general and superficial information, with a lack of geographical details regarding works financed by the program. There were no comments regarding internal management, external constraints, nor recommendations for improvements. Conversely, the 2002-2004 evaluations present more informative data about the program functioning. According to managers, the program logic is correct, since it yields economic growth, income and jobs for Northeast municipalities. As it is an IDB loan, the program is coordinated by the Ministry of Finance, but its execution is divided between the Ministry of Tourism and the States. In 2002, the bureaucracy for activities’ approvals was cited as an obstacle for the execution.

64. The physical and financial information that was supposed to be provided in the annexes was inadequate. It gives the impression that the program was not executed at all, due a total lack of indicators, budget and costs information, with 2004 as the lone exception. In the text, only output indicators are provided (number of roads km, number of airports). The 2002 and 2004 include some budget information on the text, but it is not satisfactory to allow budget decisions. There is no information that could possibly be used for learning and improvement of government-wide management.

65. Finally, there are no relevant criticisms or recommendations in any evaluation. Positive findings are consistent with TCU’s, but several risks/shortcomings that became object of TCU recommendations were not identified by program manager (e.g., change regime of tourism councils, assess cost-benefit of environmental evaluations, BNB to

⁴³ It is not clear what does this cost mean, whether it includes the transfer to the families and administrative costs, given that the transfer is around R\$ 45,00.

create performance indicators for M&E). Overall, the quality of the PPA evaluations when compared to TCU is poor.

Energia nas Pequenas Comunidades (PPA 2000-2003)/Luz Para Todos (PPA 2004-2007)

66. *Energia nas Pequenas Comunidades* provides electric energy to small rural communities using decentralized and renewable energy sources. Generally, the evaluations are too broad (except for 2002) but since the beginning it was possible to note that the program had serious problems, such as inadequate maintenance of energy generators, absence of an implementation plan, managerial turnover, etc. Unfortunately, it was only in the 2002 evaluation that the severity of the situation became explicit, in part because of the halting of program execution by TCU. Evaluations conducted in 2000 and 2001 evaluations considered the program logic correct. It was only in 2002 that manager judged it off focus due to the excessive centralization in the federal government.⁴⁴ In 2003 and 2004 the program logic changed, replaced by a more decentralized management model, which was mandated by TCU evaluation.

67. On the whole, the analysis of PPA evaluations shows that they do not provide enough information for budget decisions, mainly because management problems blocked budget execution. Likewise, there was little input to improve internal management, despite their critical content. Additionally, managers did not bring fresh ideas on how to improve program's performance. Some of the 2000 shortcomings were not solved until 2004 and when changes finally occurred, they seemed to be caused by TCU's mandatory evaluation.

68. As with other PPA evaluations, specific information on budget/execution/indicators /costs is extremely incomplete. The only PPA indicator (*taxa de localidades remotas atendidas por energia eletrica*) was listed as “*não apurado*” in all evaluations.

69. The analysis conducted by TCU raised a series of concerns regarding program implementation failures. For instance, installation of generators in places already served by electric network, lax equipment control/maintenance (including missing generators), purchase of imported equipment rather than development of national technology (one of program objectives). Similar to other evaluations, the work conducted by TCU was substantive and specific than the PPA evaluations. For example, in order to prepare the *Energia em Pequenas Comunidades* evaluation, TCU visited 11 states (Amazonas, Acre, Paraíba, Rio Grande do Norte, Rondônia, Piauí, Pernambuco, Bahia, Goiás, Minas Gerais, Rio de Janeiro) and 71 communities, interviewing beneficiaries and managers, which in part explains the capture of relevant shortcomings in the program.

⁴⁴ Due to five turnovers in the Energy Secretariat, federal managers remained two years without contact with local managers and lost control over equipment.

Programa de Saneamento Básico (2000-2003)/Saneamento Ambiental Urbano (2004-2007)

70. *Programa de Saneamento Básico* is a Ministry of Cities program that provides sanitation, water and solid waste works to small municipalities (less than 30.000 people). The five evaluations reviewed (2000-2004) convey some useful information to follow the evolution of the program, but information could be better organized. If evaluations are to serve for broader accountability purposes, besides informing the Ministry of Planning, it should bring a brief explanation about how programs operate. The evaluation could also bring the results of studies/researches contracted by the program. According to the manager, 29 researches were contracted to evaluate the program, but results were not mentioned in any of the evaluations. There is also lack of outcome indicators. Only in 2002, there is reference to IBGE census (2000), bringing coverage rates for sanitation, water and garbage collection in Brazil. The manager does not use PNAD, arguing that it is sample based and does not reflect expansion rate of sanitation. Nevertheless, PNAD has been extensively used by TCU and other important institutions for data reference.

71. Regarding the hard data brought by the annexes, it is not very useful. There is always a mention to budget execution in the text. However, this information is not consistent with the annex figures.⁴⁵ The three PPA indicators were never measured because the *Pesquisa Nacional de Saneamento Básico- PNSB* is only done every ten years. The cost annex is also of very little utility. Most of information is indicated as “not applicable”. It would be hard to take budget decisions only by reading the evaluations. It is clear that unpredictability of funds release causes serious implementation constraints. There is also reference that Congress amendments are pulverized, so it should be applied more strategically (i.e., only in poor municipalities). Nevertheless, it is not clear whether budget is enough, how it could be better programmed and better used.

72. It is possible to infer the program logic is right, given that it focuses on municipalities with the lowest HDI and worst epidemiological indicators. PSB faces several challenges that are briefly explained in the text: weak technical capacity in municipalities, absence of a government sanitation policy, potential beneficiary population is much higher than PPA target. But, these issues are not followed by analysis or ideas on how to solve them. The PPA evaluation could serve as a moment for search of relevant indicators, research papers and specialist opinions that would subsidize managers' day-to-day work. It does not mean that evaluations are self-serving. A lot of shortcomings are raised in the text. But most of them are either attributed to municipalities' weak capacity or legal issues (i.e., environmental licenses, land regularization), human resources shortage, low salaries of sanitation technicians, etc.

73. All this information is very consistent with TCU's evaluation. But unlike TCU, PSB evaluations do not bring outcome indicators. In fact, TCU successfully demonstrated the PSB positive impact using SIH/SUS data by comparing the indicators for water borne

⁴⁵ The way managers calculate inform budget execution differs from PPA/SIGPLAN format, because manager usually includes arrears and *contingenciamentos* in the analysis.

diseases (e.g., diarrhea) in municipalities served by PSB, municipalities served by other programs and municipalities not served by sanitation at all.

Educação de Jovens e Adultos (2000-2003)/Brasil Alfabetizado (2004-2007)

74. *Educação de Jovens e Adultos* (2000-2003) which later became *Brazil Alfabetizado* (2004-2007) is a Ministry of Education program that provides literacy courses to young and adult illiterates with limited years of schooling. According to PPA evaluations, national illiteracy rates decreased from 13.8% in 1999 to 11.8% in 2002, and large numbers of people were enrolled in EJA courses, which in part illustrate the effectiveness of the program. However, early evaluations (for the 2000-2003 periods), presented very little information for those not associated with the program to understand how it works. For those years, no meaningful shortcomings were identified by managers and budget information demonstrated high execution of funds.

75. In 2003, TCU performed their evaluation for EJA with case studies in six States (Alagoas, Ceará, Maranhão, Distrito Federal, São Paulo and Pará). The methodology consisted in interviews with university coordinators, secondary data analysis, direct observation and focal groups with teachers and students. In their evaluation they found serious shortcomings in the program. For instance, they discovered that after a six months course, 12.49% of students did not write, 34% wrote only words, 30% wrote only sentences and 24.05% were able to write texts. Additionally, large amount of students repeated the course, receiving the material twice and causing distortion in statistics. Finally, there was a high teacher turnover, provoking a decline in the quality and costs increase in capacity building. None of these problems were pinpointed by the managers in the four years of evaluation.

76. Given the diagnosis, TCU mandated several changes in EJA. Among the recommendations were: (a) increase in the length of the course, (b) retaining good teachers in the program and, (c) keeping a better record of students to identify the repeating ones who already have course material. Therefore, it was not by coincidence that the new PPA (2004-2007) outlined a reformulated program to tackle illiteracy, *Brasil Alfabetizado*. In order to improve quality, courses were extended to 8 months, teacher wages were readjusted to a minimum of R\$120,00 instead of R\$ 7,00 per student and the System *Brasil Alfabetizado* was implemented to keep record of partners, teachers and students. The quality of 2004 evaluation was improved over the previous years. The changes that occurred in the program seem to be a good example of how external evaluations can have impact on improving public policy.

Programa Saúde na Família (2000-2003)/ Programa de Atenção Básica em Saúde (2004-2007)

77. *Programa Saúde da Família* (2000-2003) is a program to improve access to basic health services to population through preventive care. In PPA *Brasil de Todos* (2004-2007) PSF, due to its expansion, became *Programa de Atenção Básica em Saúde*. Overall, the PSF evaluations are very informative. In the period of 2000-2001, the reports were concise and provided limited information on operations, shortcoming and

achievements. This later changed in the period of the 2002-2004, when evaluations expanded to incorporate additional information on financing, design and implementation. The 2002-2004 is also more robust regarding data provided, bringing input, output and outcome indicators such as: coverage rate of PSF, population and municipalities covered, number of teams, health agents and infant mortality rate on population served by PSF.

78. According to evaluations, PSF principles and capacity to organize health care system demonstrate that it has been able to change the health model from curative to preventive care. Moreover, the model yields important impacts on health outcomes. This diagnosis perfectly matches with the TCU's evaluation that praised the program logic by confirming the positive evolution of health indicators. Program manager was able to identify important shortcomings in the program, like high turnover and difficulty in hiring medical doctors due to the dissatisfaction with temporary contracts, lack of interest of medical schools to develop PSF curriculum, and population's unawareness of PSF model; each of these were areas identified by TCU as problematic. Conversely, other TCU findings were not captured by manager's evaluations (e.g., the excessive number of families under one team responsibility or lack of medicines in PSF units). Additionally, TCU verified that expansion was not followed by infrastructure improvements (i.e., USF physical space, human resources, and training).

79. Overall, PSF evaluations demonstrated a high level of quality. In order to improve, they could include more detailed analysis rather than the brief mentions about changes. Several PSF improvements are cited in the texts, such as *Piso de Atenção Básica*, an interesting financing scheme that could have been better explained to be used for government learning purposes. Another point that warranted further clarification was the Ministry of Health reorganization - the inclusion of *Departamento de Atenção Básica* into *Secretaria de Atenção a Saúde* and creation of *Secretaria de Gestão do Trabalho em Saúde*. Finally, possible suggestions to solve problems could be incorporated in the text.

Novo Mundo Rural- Consolidação de Assentamentos and Assentamento de Trabalhadores Rurais (2000-2003)/Assentamentos Sustentáveis para Trabalhadores Rurais (2004-2007)

80. *Novo Mundo Rural* provides infra-structure (roads, energy and clean water) and technical advice to land reform settlements and families in order to make them sustainable. PPA evaluations have been irregular and contradictory about program results. The first report (2000) was very brief but it was clear that program was going through severe problems. In 2001, the *Novo Mundo Rural* changed managers and the evaluation did not revealed changes. Conversely, the program was described, some positive results were mentioned, but only in recommendations manager cited the possibility of changing the program. The 2002 and 2003 evaluations were more informative, detecting some problems that would be later analyzed by TCU evaluation, published in 2004.

81. TCU visited six States (Pará, Ceará, Maranhão, Mato Grosso, São Paulo and Paraná) producing an unimpressive scenario of *Novo Mundo Rural*. INCRA, the federal

autarquia responsible for the program, was doing very little to improve settlements organization mostly due to its own internal disorganization. For instance, INCRA had 15 presidents in 9 years, only 25% of staff had university degree, and responsibilities for program actions were not clearly defined. Some of these problems were identified by managers, but substantial changes occurred only in 2004 when the two programs were merged to form the new PPA (2004-2007) program “*Assentamentos Sustentáveis para Trabalhadores Rurais*”.

82. Regarding the quality of data, most results are presented in the form of input and output indicators (e.g., number of people settled, funds for credit concession, etc). As it is common in other programs, PPA indicators were never measured and were considered inadequate by managers. Likewise, TCU also did not present outcome indicators either.

83. To some extent, all PPA evaluations at least mentioned *en passant* most of shortcomings tackled by TCU. Overall, evaluations blame budget unpredictability and insufficiency. There is no doubt that for a program that deals with credit concession and agriculture, lack of timely funds can be a very serious constraint. However, INCRA presents severe organizational deficiencies which depend mostly on political decisions.

Morar Melhor (2000-2003)/Habitação de Interesse Social (2004-2007)

84. *Morar melhor* is a program that transfers funds from *Caixa Econômica Federal* to States and municipalities to improve dwelling conditions for poor population (below 3 minimum salaries). The program was transformed into *Habitação de Interesse Social* for the new PPA *Brasil de Todos* (2004-2007). Overall, PPA evaluations (2000-2004) are general without analysis about how program operates and how much it contributes to solve the housing issue in Brazil. Most shortcomings identified by the manager (e.g., budget cuts, pulverized amendments that target non-poor municipalities, concentration of financial flows at the end of fiscal year) are recurrent in all years. There is no sign of improvements.

85. According to the manager, program is well defined, the target population is well characterized and solutions are adequate to solve the problem. However, program contributes very little to solve the housing deficit for poor because funds are not enough and there is low integration with other social programs. There is no evidence about the program's poor performance in any of PPA evaluations because hard data presented is very weak. Budget information is incomplete and does not allow acknowledgement of problems like arrears execution and budget unpredictability. Likewise, program indicators (*taxa de contribuição do programa para a redução do déficit habitacional* and *núcleo do déficit habitacional quantitativo de famílias com renda até 5 salários mínimos*) were never measured. Finally, manager does not mention any external evaluation of the program.

86. Despite the lack of evidence, program's poor performance, identified by PPA evaluations, is consistent with the findings from the operational evaluation which TCU undertook in 2004. Hence, manager has been critical, identifying issues like weak

information flow between the Ministry of Cities and Caixa, contracts delays/cancellation and fragmentation of actions generating weak impact on localities. However, TCU raises other problems like: (i) significant part of beneficiaries sells their houses; (ii) program has little impact on regularization of properties, etc. Clearly, TCU evaluation has a more robust methodology (direct observation in 17 municipalities, focus groups with 7 regional agencies from *Caixa*, mail questionnaire to 437 municipalities and interviews with 367 families in 15 settlements).

87. Hence, PPA evaluations present lower quality and coverage than TCU operational evaluation. Program manager could make an effort bring better data to improve internal management, such as income level of beneficiaries, social programs offered by municipalities, reasons for execution delays and improvements in health and quality of life indicators.

APPENDIX 4: THE FAILURE OF THE PPA: SOME HYPOTHESES

88. As laid out in the previous section, our assessment of the effectiveness of the PPA as an instrument to promote better management in the federal administration is not a very sanguine one. We have encountered a number of criticisms of the PPA, both in the limited local literature on the PPA and in our interviews. Critiques of the PPA can also be inferred, or put into context, from the international literature on planning and performance-management and on evaluation. To this end, some country-specific evidence on program budgeting is reviewed in Appendix 1, and on evaluation in Appendix 2.

89. The purpose of this appendix is to review and comment on these criticisms. We shall first gather together the various hypotheses suggested in the literature about potential or actual problems with the PPA. We will classify these under four headings: poor design, poor initial implementation, poor subsequent management, and hostile external conditions. We shall then propose an interpretation of how the hypotheses relate to each other and which we believe to be more important. Finally, we shall address the specific issue, within the PPA, of evaluation.

Hypotheses about the PPA

A. Hypotheses that the initial design of the PPA was deficient

90. The individual hypotheses under this heading relate to two controversial issues. First, do agencies at the center of government possess the techniques, knowledge, and foresight to plan and control in detail or must these functions be more decentralized? Second, how concretely should plan and budget be integrated and what are the conditions that allow effective integration? In respect of these issues, the PPA appears to be a more ambitious exercise than can be observed anywhere else in the world today. The uniqueness of the PPA can be resumed in three elements:

- First, every line of the budget is ascribed to one or another of around 380 programs.
- Second, these programs are meant to be units for policy rationalization (ex-ante), as well as the administrative basis for budgeting, management, and control (ex-post).
- Third, originally at least, these programs were necessarily not designed to be aligned with existing bureaucratic structures.

91. *Hypothesis A.1: “The PPA program format is a problematic organizing principle for planning and management.”* The PPA model rests on the idea that the program, as conceived under the PPA, can be an effective and dominant organizing principle for public activities. Several criticisms, often unconnected with each other, can be made of this premise:

- *Reductionism:* Coutinho (2000) and Garcia (2004) saw the PPA approach as oversimplifying the idea of the *problem* that was meant to determine the program. They argued that the PPA implicitly assumed problems to be “structured,” meaning that

variables and relationships between them are precise and quantifiable and the solution is objective. But in the real world, especially where social and political variables intervene, problems are “semi-structured,” that is, many variables and relationships are imprecise and unquantifiable and there are multiple solutions, preferences for which depend on the observer’s viewpoint and situation. When planners assume problems to be structured, they rush to unique solutions that may be wrong; that is, they may make bad policies. The PPA faces problems, as do other results-based-management systems, of measuring outputs and outcomes (Coutinho, 2000).

- *Indeterminacy*: the PPA has divided all activities into almost 400 programs that are separate and equal (in the sense of not being put into any structure or hierarchy of programs). Yet in reality, one program can have multiple objectives. It can share inputs with other programs. And one program can be nested (hierarchically) within another. This is a practical reason why a theoretically-determined program which, in principle at least, isolates programs one by one cannot constitute a unique organizing principle.⁴⁶

92. *Over-standardization*: the PPA applies the “structured” notion of program design to virtually all government activities, irrespective of their content. The question here is whether the problem-oriented definition of programs is always appropriate. Broadly speaking, different PPA finalistic programs cover routine activities (maintaining roads, making social-security payments), developmental activities (implementing a new policy such as decentralization of health services, or developing a regional pole), crisis activities (such as handling outbreaks of diseases), and policy-making activities.⁴⁷ Of course, all activities can be improved, but the problem approach (like its close cousin the logical framework) may be more appropriate, for organizational purposes at least, for developmental activities than for the other three types.

93. *Hypothesis A.2: “The PPA program format is incompatible with other organizing methods in the agencies.”* The traditional basis for government organization is a hierarchy (center of government, ministries, departments). The PPA *implicitly* sought to modify this – programs were given administrative form by their role in the budget, the appointment of managers, and an M&E system – and the program structure was meant in some ways to override the existing hierarchy.⁴⁸ The programs largely sat within ministries but did not always respect their existing departmental structures. To the extent programs assumed an important organizational function, ministries might lose some hierarchical authority to SPI and their departments might also lose some function. This potential conflict could be reduced, as a few ministries did, by modifying organizational structures to align them with PPA programs. Other ministries resisted this change (whether for technical or political reasons) by simply ignoring the program structure, except for reporting purposes. SPI and the Planning Ministry did not have the political power to force change on the ministries (beyond cosmetic adoption of the PPA format).

⁴⁶ Thus, the controversy between SPI and MEC on PPA may reflect legitimate differences.

⁴⁷ These four categories are suggested in Handy (1993).

⁴⁸ The PPA’s implicit ambition (following on the success of the management model of *Brasil em Ação* 1996-1999) seems to have been to change organization from Weberian hierarchy (a “bureaucratic culture”) to the flatter structure of a “task culture”, and from more political control to more technician control.

94. The futility of trying to change the organizational structure can be read as an implicit lesson of the international experience of program budgeting (see Appendix 1). For instance, the US Department of Defense's program budget is used for planning, but appropriations are made by organization. Kim et al (2006) take it as an axiom that two different functions related to the budget, such as planning and spending, be kept separate: "it is exceedingly difficult to subordinate the organization structure, even when the government has a program budget, because organizations actually spend the money and are responsible for results." (page 27).

95. It can be said that one instrument cannot satisfy two objectives. One instrument, the PPA program, has difficulty in serving the multiple objectives of planning/ policy making (in the sense of identifying problems-programs and rationalizing the inputs associated with them) and resource and organizational management. Had the PPA's program classification not been the basis for budget allocation, it might have corresponded to the Department of Defense's model of policy rationalization. But it was made the basis for budget allocation, which led to budget games – the manipulation of programs by both sides to gain control of budget resources.

96. Decree 5233 of 2004 tried to mitigate the conflict between program and ministry by mandating that programs be managed more within the ministry's organizational structure. The practical effect of this appears to have been limited to providing a more authoritative structure for reporting to the SPI. Ministries were still as free as they ever were not to take programs seriously.

97. *Hypothesis A.3: "The PPA management model is too simple and standardized to serve as a management tool for the agencies."* A system to manage and control almost 400 programs has to be simple and standardized to be manageable. As the responses of the officials interviewed in the ministries revealed, the system was too simple and standardized to be of use in their internal management.

98. *Hypothesis A.4: "The PPA management model is too complex to serve as an accountability mechanism between center (SPI) and the agencies."* At the same time, a system providing the center with regular information on almost 400 programs may overload the center with information. This may be the case of the PPA to judge from the annual PPA evaluation report, which provides many hundreds of pages of information on the programs but does not effectively aggregate the information. (Of course, gaps in the quantitative information complicate the job of aggregating.) Again, the foreign experience of program budgeting is relevant. Hawaii's program budget, which at its most disaggregated level had 580 programs, overloaded the system with information (Kim et al, 2006, pp. 30-31). On the other hand, Australia's 17 portfolios, replacing an earlier system of programs which provided too much information, have proven to work as an accountability mechanism.

B. Hypotheses that the initial implementation of the PPA was deficient

99. These hypotheses mostly reflect on management choices that the SPI (then SPA) initially made.

100. *Hypothesis B.1: “The SPI did not make an effort to get the agencies to ‘own’ the PPA.”* Initially, the preparatory activities for the PPA 2000-2003 preparation concentrated on program managers and deliberately bypassed the senior management of the ministries. This approach was later softened as SPI came to realize that the ministries needed to support the PPA. Even so, the PPA remained a technical exercise, and political leaders have never done much more than sign off on it. Political “ownership” of the PPA has never been high among line ministries.⁴⁹

101. *Hypothesis B.2: “PPA Programs were misidentified because the methodology was backward-looking.”* Coutinho (2000) criticizes that the PPA was prepared in a backward-looking way: problems and programs were inferred from existing Actions. This method contrasts with the US Department of Defense’s approach: “programs should be defined independently of what government is doing. As elementary as this step seems to be, it has often been among the most difficult and controversial step in program budgeting.” (Kim et al, 2006, page 28). Garcia (2004) makes the point that the PPA methodology froze processes, problems, and structures, instead of doing what strategic planning should do: allowing flexibility in identifying problems and solutions.

102. *Hypothesis B.3: “The SPI gave too few resources to the agencies to prepare for the PPA.”* Coutinho (2000) argues that resources were insufficient given the need to introduce complex new concepts. He compares the limited resources devoted to training with the more generous training sessions when Brazil introduced program budgeting in 1974 and its budget information system SIAFI in 1987.

C. Hypotheses that the SPI’s current management of the PPA is wrong

103. These hypotheses reflect on choices the SPI has made in managing the PPA, perhaps under the pressure of scarce resources or political conditions.⁵⁰

104. *Hypothesis C.1: “The SPI/MP is missing opportunities to help the agencies.”* The perceptions of the ministries are that the PPA has not been able to provide protection against budget stringencies or to act to alleviate other external constraints indicated by the evaluations. Apart from training to make the PPA work more effectively, the Planning Ministry does not appear to have provided much technical assistance to the ministries, in the area of management techniques, for instance, although the SPI has provided some help on InfraSIGs.

⁴⁹ Garcia (2004) describes the preparatory process for the 2000 PPA in the Ministry of Justice. The (early) planning process imposed by SPI (then SPA) excluded senior ministry management and ministry planners. SPI ignored the wealth of information existing inside the ministry. The ministry proposed 60 programs, but reduced this to 25 because, it was (orally) informed, SPI was targeting about 400 finalistic programs in total. SPI considerably changed these programs, altering objectives, merging programs, and excluding or merging actions. But it can be inferred that the 25 programs were more the ministry’s than SPI’s creation. Garcia also describes how discontinuity in the ministry’s leadership also undermined the PPA.

⁵⁰ There have been cuts in the SPI’s resources under the current government. It tends to have one analyst per ministry, compared to SOF, which has one coordinator and three to four analysts per ministry. SPI analysts come to their job with three months’ career training from ENAP. They acquire their knowledge of the sector from what they learn on the job (unless, exceptionally, they come with relevant sector knowledge). They are not specialists in performance management.

105. *Hypothesis C.2: “The SPI is missing opportunities to help the center of government.”* It does not appear that the SPI has been able to use the information it has gained from programs for broader purposes: evaluation results have been scarcely publicized; the SPI has played a limited role in the PPI.

D. Hypotheses that the environment external to the PPA impeded its success.

106. This largely relates to political choices.

107. *Hypothesis D.1: “The PPA is undermined because it is at odds with the reality of budget making and execution.”* A strong inference from the interviews in the ministries is that since the program approach appears to exert little influence over budget appropriations or in the solution of budget-execution problems, the PPA has lost credibility. It is clearly the case that objectives of fiscal control on the one hand and budget-prioritization mechanisms outside the PPA – notably earmarking measures and congressional amendments – on the other hand have completely frustrated the budget-planning function of the PPA.

108. *Hypothesis D.2: “The agencies can face political incentives and structural problems incompatible with the PPA.”* “Structural compatibility” with the PPA differs from one ministry to another. The interviews identified two features that can reduce the capacity of a ministry to absorb the PPA model: a high turnover of ministers and poor human resources (usually the result of the lack of new staff to inject new ways of thinking and working). We have also observed that some ministries are subject to political influences that act as counterweights to the PPA: they may be captured by particular political interests. This is so, for instance, where congressional budget amendments play a large role or where particular professions or labor groups exert a high influence.

109. *Hypothesis D.3: “The PPA does not receive the political support of this government.”* We have already discussed the relatively modest political support that the PPA enjoys from the government.

Taking a Position on the Hypotheses

110. This review of hypotheses has covered problems of the basic design of the PPA and contextual features (problems of implementation and management). Clearly some problems are more important than, or pre-empt, others. Our view is that the PPA’s standardized (universal) approach to the program format severely undermined its effectiveness and made it unattractive to the ministries. It is plausible to argue that this design problem alone was enough to dictate the poor performance of the PPA, but there have also been other important negative factors. Forcefully integrating the PPA with the budget through the programs as the links, though it has improved transparency, led managers to use PPA processes more to fight for budget resources than to improve management. Demanding constant monitoring and reporting to the SPI via SIGPlan may have also added to the ministries’ perception that the PPA existed to serve the SPI’s

bureaucratic needs rather than to help themselves improve management and implement their policies.

111. Although we first point to some of the design features of the PPA model as causes of the lack of impact on ministry behavior, effects of the external environment cannot be ignored. The PPA was introduced in a policy and administrative environment not fully conducive to performance-oriented management of public resources. The credibility of the PPA was undermined because the Ministry of Planning was unable to use the information it collected to help ministries solve problems or improve public administration, which largely lay outside its mandate or political capacity to resolve. Finally, the PPA suffered from a progressive decline in political support. This may have been bad luck in part, but it was also the consequence of the manifest failure of the PPA to contribute to government performance.

The PPA model

112. The passage of time suggests two fundamental weaknesses of the PPA model, the insistence on a “pure” program format and a choice of universality over selectivity. The two weaknesses together remind us forcibly of the fallacies of planning. (see Box 6 on problems of planning in the private sector).

Box 1: The Sins and Fallacies of Strategic Planning in the Private Sector

Strategic planning – in the sense of centralized planning using formal techniques – is well past its highpoint of popularity with private firms and business schools. Wilson (1994, quoted in Mintzberg et al, 2000) lists seven capital sins of strategic planning:

- specialized planning units usurped management;
- elaborate process dominated;
- as executives were excluded, they did not execute the plan strategy/strategy did not guide action;
- planning concentrated more on mergers & acquisitions than basic business;
- the planning process did not throw up real strategic options – the first viable option was chosen, at the expense of examining alternatives;
- planning neglected organizational and cultural requirements of the strategy, focusing more on the external than internal environment; and
- planning tended to result in a single path, rather than options/ scenarios, making plans vulnerable to surprises.

Mintzberg et al (2003) identifies three fallacies of strategic planning:

- The fallacy of predetermination: you cannot predict the future.
- The fallacy of detachment: you cannot pass hard operational data up the chain and make the top of the organization omniscient. These data miss out important non-quantifiable elements, lose out through aggregation, arrive too late, and are often unreliable.
- The fallacy of formalization: formalization discourages creativity and synthesis.

Many of these points are relevant to the PPA. The government may well have been induced by its frustration at the weak performance of many ministries to step in to promote performance (as it

later did, yet again, with the PPI). But the experience of the private sector questions the capacity of a technical exercise at the center of an organization to tell the producing departments what to do in detail. This is because the center cannot know enough about the future or about how the departments really do their business. As a result, managers at the center and in the departments tend to ignore the exercise.

113. The ambition of the “pure” approach was to put *all* government actions within a standard program format *and* to make programs the central logic for organizations. The idea makes intuitive sense, and the program format has undoubted virtues such as transparency and an emphasis on outputs/outcomes. By the standards of other reforming countries, this was a high ambition indeed. But programs cannot always be unambiguously identified, while the program logic may be at odds with organizational logic and with budget incentives – in particular, this logic may have less sense for routine than for developmental activities.

114. Moreover, universality ignores differences between organizations in terms of their objectives and technical capacities; assumes that better performance will come about through a standard formula (i.e. “one size fits all”), rather than one that differentiates between objectives and between capacities; spreads resources too thin; fails to signal true priorities of the government (and thus also lead to less support from the Presidency itself); and creates bureaucracy without results. Therefore, we now believe that trying to apply a standardized approach to improving management in different types of activity prove ineffectual.

115. It is therefore not surprising that the PPA, in particular the program format, has come to face substantial resistance. Sectors have resisted this format in order to preserve budget flexibility (e.g., by trying to define as a program an umbrella set of actions, thus rendering the “program” mere unit of budget classification and appropriation rather than a unit of management) and/or to preserve their existing organizational structures, whether out of conservatism or because the program logic does not work as organization logic.

116. Linking the PPA to the budget (or rather making the two basically identical in form) creates some benefits as well as perverse incentives. On the benefit side, the PPA’s program format and its adoption as the structure of budget appropriation have significantly increased fiscal transparency. Now oversight bodies such as Congress and the TCU can monitor budget execution by program, and reasonably detailed qualitative information about each program (beyond its title) is available for those who are interested in understanding what each of them does. The TCU has begun to conduct its annual audits by program, and started a credible program of selective program evaluations (although the TCU evaluations do not limit themselves to PPA programs as the unit of analysis). Among the most positive consequences we came across during our interviews were the cases where the ministries were mandated to improve their program and overall management by the TCU after its audit found deficiencies in the ministries’ program management.

117. On the negative side, however, the use of the PPA program classification as a structure of budget appropriation gives ministries incentives to aggregate programs into larger “umbrella” programs so as to retain greater flexibility during budget execution. This in fact seems to be what happened in those large social sector ministries. Greater flexibility in budget execution makes sense from the point of view of the executing agencies, and is the general trend in performance-based budgeting among OECD countries. But, it can have a detrimental effect on the PPA to the extent it removes one of the few tangible benefits it has brought about, transparency. We also imagine that this sort of “opportunistic gaming” where SOF/SPI and the line ministries negotiate program definition with different motives – SPI wanting to preserve the conceptual purity of what a program should be, SOF trying to limit line ministries’ spending increase, and line ministries trying to lobby for more budget – would be unhealthy for promoting performance orientations within the government. In the least, it would foment cynical attitudes about programs among line ministries.

“Hostile” political environment and the absence of complementary policies

118. Aside from the inherent limitations in the original design of the “model,” the lack of political support has been harmful to the further development of the PPA under the current administration. The PPA’s loss of political support is partly the result of bad luck, partly a consequence of PPA-specific failures. The PPA lost support under the new Administration in 2003, partly because of the incoming government’s distrust of the bureaucracy and different priorities by the new minister of planning who appeared to be more interested in economic development issues via industrial policy, and later on, via development of a new law for public-private partnerships than public management reforms. One might also infer that the lack of political support arose from the PPA’s inability to deliver. This resulted in the transfer of initiatives to other parts of the government – monitoring of key programs to Casa Civil, the leadership role in the quality of public spending agenda to the National Treasury, strategic thinking to “Brasil em 3 Tempos”, and so on. The failing credibility of the PPA has also meant a loss of support from the ministries, the Ministry of Finance, and SOF.

Changes in the Political Context

119. When the new Administration came to power in 2003, the Ministry of Planning suffered a considerable loss of political standing. The new team arrived with a high level of distrust of the administrative machine, which it associated with corruption and loss of political control (e.g., “government-by-technocrats”), and a high priority was to take control of the bureaucracy (rather than to let the public managers manage for results). Among the three central agencies, the Casa Civil, the Ministry of Finance, and the Ministry of Planning, the first two always enjoyed greater political influence than the latter. During the current administration, however, the relative balance of power shifted further away from the Ministry of Planning to the Casa Civil, which began to assert itself more on government management issues and initiated its own monitoring of priority activities through the new unit, *Subchefia de Acompanhamento e Monitoramento*, and the Ministry of Finance, which, no longer content to be the guardian of aggregate fiscal discipline, began to involve itself in expenditure prioritization decisions, a traditional

purview of the Ministry of Planning. The new posturing by the Casa Civil and the Ministry of Finance not only undermined the Ministry of Planning's traditional authority and statutory responsibilities, but also created some confusion as to who was responsible for what among the central agencies, and caused some tension among them.

120. As part of this shift, the PPA in 2003 lost the small capacity it had – through cash-flow control – to prioritize budgetary spending. Meanwhile, a system to monitor and manage the President's strategic objectives (*Sistema de Metas Presidenciais*) was set up in the Casa Civil (Box 1). The budget office (SOF/MP), an ally of SPI and the PPA during 2000-03, has moved towards the Ministry of Finance, and this has left the planning office (SPI) in greater isolation, while the Secretariat of Management (SEGES), another natural ally of SPI, at least potentially, was busy developing its own agenda of a government-wide public management reform.⁵¹ As a telling illustration of the assertiveness of the Ministry of Finance on micro-expenditure allocation issues, the initial development and coordination of the PPI initiative was led by the National Treasury Secretariat, with some tension with the Ministry of Planning at the beginning.

Box 2: *Sistema de Metas Presidenciais*

In 2004, the government came to a decision to formally make the Casa Civil the manager and coordinator of government actions, and the Subchefia de Articulação e Monitoramento (SAM) was created. The idea was of an implementing/problem-solving function, rather than a longer-term-strategy function. (This helps explain the decision almost made in 2003 to move Secretariat of Management (SEGES) from the Ministry of Planning to the Casa Civil. In the end, the government decided to leave public administration matters – such as human-resources management and the tackling of generic public-sector problems – in the Planning Ministry, and thus de facto accorded them lower political priorities) Also, the idea was to take a selective approach, not cover the whole of government (as the PPA did), and to monitor rather than evaluate. A small unit within the SAM runs the Sistema de Metas Presidenciais (SMP).

Around 80 Goals were selected for the SMP. These were largely (but not entirely) winnowed down from the list of PPA programs. Those Goals with budget allocations map reasonably well with PPA programs (though some Goals represent an amalgam of programs). There are also some Goals, such as regional trade integration, which are policy goals without budgetary implications. Our impression is that the Goals are as much the aggregations of bureaucratic/expert views as the stated political preferences of the President

The SMP is designed to provide rapid information to the President about implementation. It also provides monthly, six-monthly, and annual reports to inform the public about progress on the Goals.

Each Goal has a Manager who reports to a Monitor in SAI. Where SAI identifies an implementation problem, it first seeks to solve it in-house, failing this pushes it up to the Minister of the Casa Civil, failing this pushes it up to the President. The SAI's experience (like that of SPI) is that agencies differ in the quality of information they can provide and the extent of the cooperation they offer in problem-solving.

⁵¹ SEGES never developed a full reform plan before the secretary who was leading the charge left the Ministry because disagreements with the Minister.

We have no independent confirmation of the effectiveness of the SMP in problem solving, but to the extent that it is effective, it is a management-information system that does what SIGPlan/SPI might in part be expected to do, hence undermines the PPA's legitimacy.

121. The decline in high-level political support for the PPA went hand in hand with the absence of a clear public management reform agenda by the federal government. As the Bank's previous report emphasized, the PPA, as a planning and performance management device, cannot operate in isolation of the environment of the public administration. The unreformed federal administration was not the most propitious ground for incremental progress in performance-oriented initiatives.

122. The same political context that led to the dwindling political support for the PPA also explains the lack of further development in complementary policies that were needed to support and strengthen the PPA. The most important failure in this regard has been in the area of expenditure management where the combination of extremely high levels of earmarking and policy emphasis on fiscal balance have virtually eliminated the PPA's capacity to influence budget prioritization. A government-wide plan without the ability to establish and protect policy and spending priorities is a handicapped instrument from its inception.

123. Effectiveness and efficiency of program implementation also depend on other public management functions such as human resource management and government procurement. In the former, the government has recently undertaken no major initiative to make human resource management more performance-oriented. In procurement, the government has adopted some promising measures such as an expanded use of *pregão eletrônico*, and rationalization of government purchases (at least in the first year or two). But the cumbersome process involved in government procurement remains a constant complaint of program managers, and a major overhaul seems beyond the horizon.

The role of SPI/MP

124. The Ministry of Planning in general, and the SPI in particular, have missed opportunities, on several fronts. The SPI has not been able to act, apparently for lack of appropriate authority and/or resources, to respond to implementation difficulties faced by programs and sectors. Similarly, the MPOG (primarily SPI and SEGES) has not been able to put to use the management information collected during the evaluation process. For example, the SPI was not able to respond to the government's information needs associated with PPI. Had it been a center of knowledge and of strategic-thinking, it could have informed the government about the poor state of readiness of the initial set of PPI projects. The PPA's ability to protect resource allocations to priority programs, weak to begin with, has weakened further with the decision to abandon the cash-flow control. SPI could have done more to publicize the PPA or evaluation results or to get feedback about the PPA from the ministries, although we would never know whether such a pro-active "marketing" would have been sufficient to maintain a needed level of government support for the PPA.

In our view, SPI has missed the opportunity – because it chose the road of universality rather than selectivity – to improve specific elements of the sectors’ capacity for performance management, notably in the areas of information, management (and indicators), evaluation capacity, and planning capacity. In hindsight, it seems to us that SPI has tried to create a system before the sectors had sufficient capacity for this, and in so doing, neglected to pursue an alternative path of building on different ministries’ strengths and weaknesses in a more tailored approach that would leverage ministries’ own interests and commitments to a higher degree.