For Official Use **GOV/PGC/REG(2004)6**

Organisation de Coopération et de Développement Economiques
Organisation for Economic Co-operation and Development

15-Oct-2004

English - Or. English

PUBLIC GOVERNANCE AND TERRITORIAL DEVELOPMENT DIRECTORATE
PUBLIC GOVERNANCE COMMITTEE

GOV/PGC/REG(2004)6
For Official Use

Cancels & replaces the same document of 16 September 2004

Working Party on Regulatory Management and Reform

REGULATORY PERFORMANCE: EX-POST EVALUATION OF REGULATORY TOOLS AND INSTITUTIONS

Draft report by the Secretariat

27-28 September 2004
OECD Headquarters, Paris

For further information regarding this document please contact Peter Ladegaard (tel: +33 1 45 24 19 06, email: peter.ladegaard@oecd.org).

JT00171883

English - Or. English

REGULATORY PERFORMANCE:
EX POST EVALUATION OF REGULATORY TOOLS AND INSTITUTIONS

TABLE OF CONTENTS

REGULATORY PERFORMANCE: *EX POST* EVALUATION OF REGULATORY TOOLS AND INSTITUTIONS 3

EXECUTIVE SUMMARY 5

1. INTRODUCTION 7

1.1. Why evaluate regulatory policies? 7
1.2. How can regulatory tools and institutions be evaluated? 9
1.3. Organisation of report 9

2. EVALUATING REGULATORY POLICIES AND PUBLIC SECTOR PERFORMANCE: CONTEXT AND CONSTRAINTS 11

2.1. Development of evaluation practices and key drivers 11
2.2. Form of evaluations 14
2.3. Choosing the appropriate evaluator 15
2.4. Methodological issues 17

3. A CONCEPTUAL FRAMEWORK FOR *EX POST* EVALUATION OF REGULATORY TOOLS AND INSTITUTIONS 20

3.1. Compliance tests (process) 21
3.2. Performance tests (output) 21
3.3. Function tests (outcome) 23
3.4. Conclusion 24

4. REGULATORY IMPACT ANALYSIS 26

4.1. Introduction 26
4.2. Compliance tests 27
4.3. Performance tests 30
4.3.1. Models and Methodologies 34
4.4. Function tests 35
4.5. Conclusion 38

5. CONSULTATION MECHANISMS 41

5.1. Introduction 41
5.2. Compliance tests 42
5.3. Performance tests 43

5.4. Function tests 47
5.5 Conclusion 50

6. REGULATORY OVERSIGHT BODIES 52

6.1. Introduction 52
6.2. Compliance tests 54
6.3. Performance tests 54
6.4. Function tests 56
6.5. Conclusion 58

7. TRENDS AND CHALLENGES: TOWARDS MORE EFFECTIVE MONITORING OF REGULATORY PERFORMANCE 59

7.1. Political commitment, institutional and cultural challenges 59
7.2. Effective and timely monitoring of regulatory performance 61

8. CONCLUSIONS 65

BIBLIOGRAPHY 71

ANNEX 1: THE OECD REFERENCE CHECKLIST FOR REGULATORY DECISION-MAKING 73

Tables

Table 1. Who Evaulates? Advantages and Disadvantages of Different Approaches 16
Table 2. Positions of Evaluators and Their Relation with the Evaluated Regulatory Tool or Institution 17
Table 3. Matrix of evaluative tests: roles of regulatory oversight bodies & evaluation types 58
Table 4. Classification and examples of tests to evaluate regulatory tools and institutions 67

Boxes

Box 1. Definitions of Regulatory Policies, Regulatory Quality and Evaluation 8
Box 2. Examples of Policies and Strategies to Evaluate Regulatory Tools and Institutions 14
Box 3. Common Characteristics of RIA 27
Box 4. A Draft Checklist for Evaluation of Regulatory Quality Tools and Institutions 70
Box 5. The OECD Reference Checklist for Regulatory Decision Making 73

EXECUTIVE SUMMARY

1. Regulatory reforms need timely, high-quality economic and technical analysis capable of speaking to both general and technical policy audiences, to assist in decision-making and to assure the integrity of the regulatory process. But how do we know when the applied regulatory tools actually achieve these goals? Do regulatory policies deliver high quality regulation and better regulatory results?

2. Knowing more about the efficiency and effectiveness of the applied regulatory tools can support policy-makers in improving regulatory outcomes and reducing the risk of regulatory failures. Evaluation of regulatory policies also provides a mechanism to assess costs and benefits of the significant resources invested in regulatory management systems established in most OECD countries.

3. This report looks at the evaluation of regulatory tools and institutions from three perspectives. First, it provides an overview of OECD Countries' practices with *ex post* evaluation of regulatory tools and institutions. It focuses specifically on the two most important regulatory tools – RIA and consultation mechanisms – and on one of the most important and increasingly applied regulatory institutions, the regulatory review body. Second, it develops a conceptual framework to assist in clarifying the range of potential evaluation tests. And third, it proposes a draft checklist to guide government strategies to evaluate regulatory tools and institutions.

4. The report concludes that OECD Countries generally have only limited experiences with *ex post* evaluation of regulatory tools and institutions. Evaluation activities are recent and often ad-hoc, generating very little quantitative or comparable data.

5. The conceptual framework for potential evaluation tests suggests a three-part taxonomy of evaluation tests, differentiating between compliance (process), performance (output) and function (outcome) tests:

- ***Compliance tests*** evaluate formal compliance with the procedural requirements of the regulatory quality tool or institution, as set out in laws, policies or guidelines as appropriate.

- ***Performance tests*** measure the quality of the analysis undertaken, going beyond the question of formal compliance with procedural requirements.

- ***Function tests*** evaluate the actual effect of the regulatory tool or institution on the quality of the regulatory outcome.

6. The report also identifies a range of specific compliance, performance and function tests. It argues that there should be a correlation between the kinds of tests employed and the sophistication/experience of the regulatory framework in which the country finds itself. Relatively simple compliance tests should be favoured in the early stages of implementation of a regulatory tool or institution. Performance tests should increasingly be favoured as expertise in the application of the tests is developed and where there is a greater concern with the quality of their application, rather than simply to ensure that they are applied at all. And outcome tests should be used to test whether a fully functioning regulatory policy tool is, in fact, having the predicted effects in increasing regulatory quality.

7. Country practices, however, show no clear pattern of emphasis in terms of movement between the different types of tests over time. Countries with limited experience in implementing regulatory quality tools – or of evaluations of them – sometimes attempt to implement much more advanced tests, without first obtaining a sound understanding of basic compliance issues. The analytical framework is considered valuable because it can be used by regulatory reform authorities as the basis for organising evaluations and timing the application of the different tests and approaches.

8. Countries face a number of challenges in the pursuit of (better) evaluation of regulatory tools and institutions, most notably related to data, institutional constraints, and cultural barriers. The report suggests that many of these challenges can be addressed through changes to the regulatory process itself, not least by making appropriate revisions to existing RIA procedues. Paradoxically, this increased emphasis on *ex post* evaluation puts more pressure on the *ex ante* tasks of policy design and analysis.

9. The report also includes a draft checklist, which identifies fundamental considerations of *ex post* evaluations of regulatory tools and institutions in the form of twelve questions.

1. INTRODUCTION

1.1. Why evaluate regulatory policies?

10. In a context of reduced scope of traditional macro-economic tools, constraints on budgetary spending, and substantial and complex demands from citizens, the role of regulation as a policy tool has progressively increased since the 1960s and 1970s. As the scope and scale of regulatory interventions continues to expand, the costs of poor quality regulation have also risen apace. Thus, it has become crucial to put in place tools to assure regulatory quality.

11. The rise and development of regulatory policies (see Box 1) over the last two decades is proof to governments' responses to these challenges, and illustrates the growing awareness and priority of regulatory quality.[1] Increasingly, regulatory policies have moved closer to the centre of many governments' policy agendas and their implementation has included adopting sets of tools designed to improve the quality of government regulation and the building of an institutional framework to apply them. Substantial resources are being allocated - directly or indirectly - to the design, application and improvement of rulemaking procedures.

12. As governments progress in the development of regulatory policies, there is growing attention paid to evaluate the outcomes and assessing the performance of regulatory tools and institutions. The emerging interest reflects three inter-related developments:

- First, policy-makers involved in regulatory policies are being held *accountable* for the significant economic resources as well as the political capital invested in regulatory management systems now established in most OECD countries.

- Second, there is a growing interest in exploring how regulatory policies can be more *evidence-based* and supported by empirical findings. More evidence-based approaches to the assessment of regulatory quality allows for a review of the effectiveness of policy tools used in practice, for a review of their performance and for improving the design and implementation of the policy.

- Third, the move toward *ex post* evaluation is part of the *progressive development* of regulatory policies, complementing the current dominant focus on *ex ante* evaluation, and aligning regulatory evaluation with the evaluation of other government policies and activities.

13. The opportunities to conduct *ex post* evaluations of regulatory policies and to respond to their results with appropriate adjustments are substantially greater than in past years. This is the result of the accumulation of policy learning in relation to these instruments. There is emerging consensus regarding good practices, or even "best practices" in relation to particular tools and institutions and their application.

[1] The 1995 OECD Recommendation on Regulatory Quality represented, for the first time, an internationally accepted set of principles on ensuring regulatory quality was adopted. However, at the time of the Recommendation, only a minority of Member countries had put in place formal policies to ensure that such principles could be implemented systematically. By 2000, 24 of 30 OECD Member countries had adopted regulatory policies, yet in at least ten of those countries, the policy had been introduced during the previous five years. OECD (2002).

This provides opportunities to "benchmark" existing practices, both in terms of the content of particular tools and in terms of the quality of their practical implementation.

14. The evaluation of regulatory policies should also be seen as an important issue within the broader governance agenda, as well as in the broader context of the evaluation of government policy and programmes. There are clear and substantial similarities in terms of both the underlying objectives of evaluation and the methodological and practical issues and difficulties.

15. The purpose of the evaluation of regulatory policies is to improve the performance of regulatory quality tools and institutions – measured in terms of their ultimate goal of increasing the effectiveness and efficiency of regulation over time. Thus, the evaluation is based on improving the performance of government and its accountability for that performance to its citizens. This broader context provides an additional spur to the evaluation of regulatory policy tools, since it is increasingly necessary for the proponents of particular government activities to be able to justify their importance and functions on the basis of objective data if these activities are to continue to be supported. Developing a sound understanding of the practical performance of regulatory quality tools and institutions through rigorous evaluation can therefore assist policy-makers in making the case for continuing to devote resources and effort to these ends and provide the basis for further entrenching regulatory policy within the core of government and expanding the scope and depth of these instruments. The evaluation of regulatory tools and institutions is crucial to underpinning their continued existence, maximising their performance and, as a result, maximising the effectiveness and efficiency of the regulatory policy itself.

16. From an OECD-wide perspective, information sharing on the performance of different regulatory quality tools, the determinants of that performance and the interactions between tools and institutions can lead to improving cross-country learning and the avoidance of unnecessary policy failures. Thus, the rate of policy learning can be increased – possibly substantially. This is clearly an important consideration given the relatively early stage of development of regulatory quality tools and the consequent substantial gaps in policy-makers' knowledge of the nature of the tools.

Box 1. Definitions

Regulation refers to the diverse set of instruments by which governments set requirements on enterprises and citizens. Regulations fall into three categories: Economic regulations intervene directly in market decisions such as pricing, competition, market entry, or exit; Social regulations protect public interests such as health, safety, the environment, and social cohesion; and Administrative regulations are paperwork and administrative formalities through which governments collect information and intervene in individual economic decisions.

Regulatory tools and institutions refer to the mechanisms by which governments promote regulatory quality, consistent with their underlying regulatory policies. Examples of regulatory tools include regulatory impact analysis (RIA), consultation and communication mechanisms, simplification measures such as time-limits for decision-making, sunsetting and automatic review clauses. Regulatory institutions include central regulatory quality oversight units, external committees (established by government with the purpose to promote, propose or implement various regulatory quality measures), and independent regulators.

Regulatory policies are policies designed to maximize the efficiency, effectiveness, transparency and accountability of regulation based on an integrated and rational approach to the application of regulatory tools and institutions. Regulatory policies focus on creating the optimal framework for the process of producing and reviewing regulations, rather than on the material content of regulations per se.

Regulatory quality refers to the extent to which a regulatory system pursues its underlying objectives. These objectives involve the specific policy objectives which the regulatory tool is being employed to pursue and the efficiency with which those objectives are achieved, as well as governance based objectives including transparency and accountability. To decide whether a system of regulation is of high quality, or in need of reform, it is necessary to be clear about the benchmarks that are relevant in such an evaluation. The OECD's Reference Checklist for

Regulatory Decision-Making sets out ten general criteria and principles for regulatory quality, which have been widely applied by OECD Member Countries in designing and implementing regulatory procedures. (See Annex 1)

Evaluations are systematic and analytical assessments of important aspects of a policy, programme, organization, regulation or other matter. The role of evaluation is to improve the information base for future decision-making. Effective regulation thus creates the conditions for a "virtuous circle" of policy analysis and improvement to develop through the implementation of effective "feedback loops", by which the results of the evaluation are translated into policy changes. A distinction can be made between *ex ante* evaluations performed before the implementation of the object of evaluation, and *ex post* evaluations which takes place after completion. However, it is important to recognise that, in both cases, even evaluations based on rigorous methods rely significantly on subjective judgements.

1.2. How can regulatory tools and institutions be evaluated?

17. Since the underlying purpose of regulatory quality tools and institutions is to enhance the quality of the regulatory structure, it follows that the best means of evaluating them is to look directly at regulatory quality. However, while this is clearly a theoretically optimal approach, the practical problems of making an assessment of aggregate regulatory quality – or even, less ambitiously, of the quality of the "flow" of new regulation (as distinct from the overall stock of regulation) - is widely recognised by regulatory reformers. While some few attempts have been made, particularly in the United States, to measure the aggregate costs and benefits of regulation, the results are subject to substantial uncertainty.[2]

18. The approach taken in this report is, therefore, focused on the identification and development of a range of more specific and "technical" tests to evaluate regulatory tools and institutions. Most of these tests are, when considered individually, limited in their usefulness. That is, they can be expected to provide insight into one or more specific aspects of the use of a regulatory quality tool, but are unlikely to allow an analysis of its overall functioning. However the nature of the insights is that they tend to be more detailed and specific, and thus likely to be more suitable to the task of guiding the optimisation of specific aspects of a regulatory quality tool or institution. Moreover, such tools are in many cases complementary in nature, suggesting that the challenge for policy-makers is to design appropriate and feasible combinations of the tools that will, taken together, achieve the evaluative objectives.

19. A further advantage of the narrower and more "technical" approaches to evaluating regulatory quality tools proposed in this report is that they are more likely to allow some of the links between the application of the tools and improvements in resulting regulatory quality to be understood and highlighted. As noted above, this is an essential element in driving improvements in the design of these tools over time.

20. The project has been carried out under the auspices of OECD's Public Governance's Committee and its Working Party for Regulatory Management and Reform. Preliminary findings and results of a survey carried in mid 2003 were presented at an Expert Meeting in OECD Headquarters in Paris on 22 September 2003. The current report builds on this earlier work and also draws on a number of commissioned papers by experts in the field.

1.3. Organisation of report

21. The report is set out as follows:

[2] It can be noted, however, that other forthcoming projects of the OECD's Public Governance Directorate – such as the continued development of indicators of regulatory quality and a Red Tape Scoreboard to measure administrative burdens, are taking this more "aggregative" and empirically based approach to evaluating regulatory quality further.

- ***Chapter 2*** discusses the issue of evaluation in the public sector context in general terms and examines how this relates to governments' specific experiences with *ex post* evaluation of regulatory tools and institutions. It identifies some of the main drivers of evaluation activity, considers the question of who should conduct evaluation and why and discusses some basic methodological issues relating to evaluation generally.

- ***Chapter 3*** considers the evaluation of regulatory quality tools and institutions at a theoretical level. It proposes a three part taxonomy of evaluative tests, comprising compliance tests, performance tests and functional tests. This three-part taxonomy is considered likely to be applicable to a broad range of regulatory quality tools and institutions and so forms a large part of the theoretical framework for the report.

- ***Chapter 4*** applies the theoretical framework to the evaluation of Regulatory Impact Analysis (RIA). The chapter mixes theoretical discussion of different possible evaluative tests for RIA with practical discussion of the application of such tests in various OECD Member countries. It brings forward key conclusions of these tests as applied in practice and considers their importance in terms of the theoretical framework.

- ***Chapter 5*** follows a similar approach to the evaluation of consultation and communications policies. The three-part taxonomy is found to be broadly applicable to these regulatory policy tools.

- ***Chapter 6*** considers means of evaluating central regulatory oversight units. It finds that this issue is complicated by the threefold role of such units, embracing advisory, gate-keeper and advocacy functions. Nonetheless the three-part taxonomy of compliance, performance and function tests is found to be applicable to the evaluation of these units.

- ***Chapter 7*** presents an overview of trends and challenges in the evaluation of regulatory quality tools, as well as discussing some of the broader-scale attempts to assess regulatory quality directly.

- ***Chapter 8*** presents general conclusions based on the analysis contained in the report.

2. EVALUATING REGULATORY POLICIES AND PUBLIC SECTOR PERFORMANCE: CONTEXT AND CONSTRAINTS

22. This chapter presents a general discussion of issues in public policy evaluation in order to provide a broader context for the following discussion of the specific issue: governments' efforts to evaluate regulatory tools and institutions. The discussion draws on the past work of the Public Governance Directorate in the area of public policy evaluation, as well as the findings of a survey of member country practices in relation to evaluating regulatory quality tools and institutions carried out as part of this report. A key perspective is to draw out the conceptual links and the key distinctions between public policy evaluation generally and practice in the specific area of regulatory policy tools and institutions.

2.1. Development of evaluation practices and key drivers

23. Governments adopt a range of approaches to measuring and improving their performance in implementing public policy in pursuit of societal objectives. As an integral part of government processes, this is achieved through strategic management, programme evaluation, result oriented budgeting and management, and performance reporting and auditing. The formalisation of goals and measures in the governmental process assists policy alignment, performance management and accountability.

24. As the concept of evaluation is becoming increasingly recognised, virtually any type of feedback or inquiry is increasingly likely to be referred to as evaluation. A number of terms, such as review, follow-up, monitoring, audit, scrutiny, assessment are often used to refer to evaluative activities. Some experts have expressed concern about the possibility of the concept of evaluation losing its specific meaning as a result of this broadening of the definition of what can constitute evaluation. Furthermore, there are other feedback mechanisms, in addition to evaluation, that can be used to improve decision-making. The term "evaluation", in the context of this report, is used to refer to systematic and analytical assessments of important aspects of a government activity (here: regulatory tools and institutions) and its value, with a view to creating or enhancing policy feedback – that is, enhancing the future performance of the activity being evaluated.

25. Three stages of development of public sector evaluation activities have been distinguished by theorists. The first wave of evaluation, occurring in the 1960s and 1970s, was largely linked to the activist approach of numerous liberal and social-democratic governments during the period in launching a range of new programmes to solve social problems, with favourable fiscal conditions and the increased status and supply of social science knowledge contributing to substantial efforts being made in this direction. Within this context, a need for more sophisticated programme planning processes was recognised to ensure the quality and utility of these new – and often highly resource intensive – initiatives. Programme managers increasingly turned to evaluation to provide a feedback mechanism, validating the newly implemented programmes and providing the basis for achieving further improvements over time.

26. The second wave of interest in evaluation occurred in the 1980s, although it is considered less striking and rapid in its development. This "second wave" was stimulated by predominantly conservative governments attempting to curb public programmes given fiscal constraints. Evaluation was thought to be useful in reconsidering the justification of policies and rationalising resource allocation within the budget. That is, support for evaluation was, in many cases, based on the presumption that there were significant programmes that could not withstand rigorous scrutiny of their contribution to the social objectives that

formed their identified *raison d'être.* The over-riding presumption during this period was that, where evaluation identified such programmes, the policy response would be to abolish them, thus tending to reduce government expenditure levels. Thus, evaluation programmes could be seen as being to some degree captured by an ideological perspective during this time. Ministries of Finance and Audit Offices were active in developing evaluation activities in this period.

27. In recent years, the focus of policy evaluation has been shifting towards a more comprehensive approach which aims to assess efficiency as well as the quality of governance. Policy evaluation is increasingly becoming an integral part of most OECD Governments' public sector reforms, although with important differences in emphasis, approach and focus. Compared with previous evaluation efforts there are now more realistic expectations, more widespread acceptance of less rigorous methodologies and greater understanding of issues concerning the utilisation of the results of evaluation within organisations. There is also a strong emphasis on more systematic, outcome-oriented evaluation with linkages to the budget process.

28. As is suggested above, making use of the findings of evaluations is not an easy task. The history of evaluation may be characterised as one of unfulfilled promises. Therefore, it is unsurprising that doubts are raised in many quarters about its overall usefulness. Some see evaluation as a management fad that creates bureaucracy but delivers few results. It is sometimes seen as inherently too theoretical to work in practice and as affecting issues of marginal significance rather than major policy choices. Others fear increased control or do not wish to be held accountable. Those primarily interested in the continuation of programmes may feel threatened by evaluation. The above observations regarding the "second wave" of evaluation, where it was specifically deployed in many cases as a means of identifying programmes for elimination suggest there are specific realities underlying such fears. At the same time, the notion of evaluation necessarily includes the possibility of a judgement that a programme is failing and the option of discontinuing it, rather than seeking to reform or improve it. Thus, the tendency for programme managers to fear evaluation can, to some degree, be expected to remain a constant.

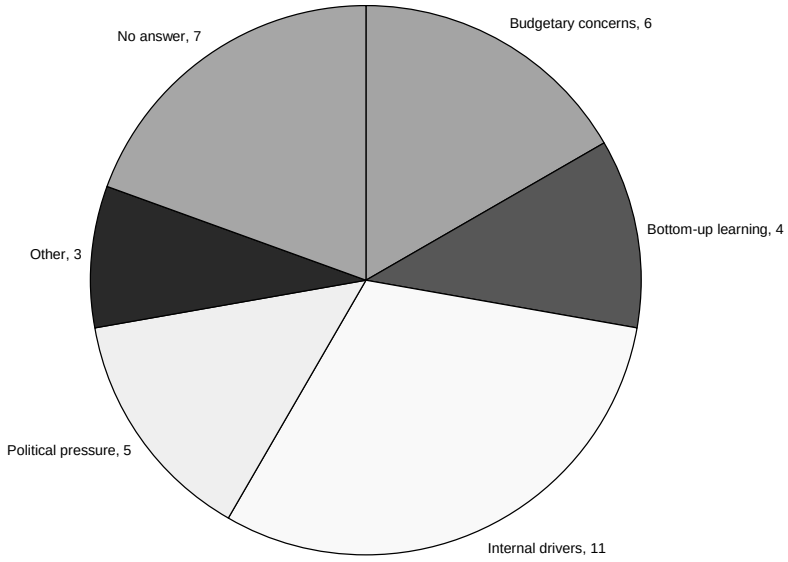
29. Questions can be raised about the value of a distinct evaluation process in an environment where performance is continuously measured and even contracted for – whether via external service provision contracts or through the use of performance indicators within the public sector itself, with their common linkages to the budget process. In this context, it can equally be argued that evaluation is, in fact, an inevitable corollary of recent moves toward more accountable public sectors, based on responsibility for achieving identified objectives and that the challenge is to integrate a more "strategic" approach – as traditionally taken in the evaluation context – into a culture now characterised by constant reporting against short-term performance indicators. Evaluation can fulfil an important role when properly used and integrated with the overall performance management framework. It can improve the efficiency and effectiveness of the public sector and hence strengthen the basis for public sector activities.

30. Evaluation mechanisms are typically resource intensive and this characteristic tends inevitably to raise questions about the cost-effectiveness of the mechanisms themselves. That is, if government programmes and policies are to be subject to benefit/cost based analysis and benchmarking, it is *a fortiori* necessary for benefit/cost analysis to be applied to the analytical process itself. As well as the direct resource costs of the evaluation activity itself, it is often pointed out that evaluations can be disruptive for to the day to day operations of the agencies or programmes being evaluated and thus can themselves have a negative impact on the implementation of policies. Put another way, the resources involved in evaluation are not solely those of the evaluating body, but also include resources diverted from programme activity by programme managers and staff to assist in providing data and interpretations to evaluators.

31. How are these general characteristics and trends of public policy evaluation reflected in governments' efforts to evaluate the performance of regulatory tools and institutions? First, it is necessary

to emphasise that there very little tradition and experience with the evaluation of regulatory policies, in particular *ex post* evaluation of regulatory tools and institutions. Regulatory policy itself is a recently developed and still developing policy area, so that the lack of evaluative activity is to a substantial extent to be expected: in most countries, the focus of regulatory reformers remains on the implementation of these tools and institutions and the need to embed them as part of the core regulatory processes and institutional structures of the country, rather than the "next stage" question of whether the particular variants implemented have been effective in meeting their goals. However, this situation can be expected to change quite quickly, since it is precisely when a new and formerly untried policy has been recently implemented that the need to determine its performance in practice is likely to be strongest. From an efficiency perspective, the will to adjust policies and institutional arrangements is likely to be strongest in the early stages, before they have become a part of the "mainstream" structure of government.

32. When asked to identify the drivers of governments' efforts to evaluate regulatory tools and institutions, the majority of surveyed countries indicated that their evaluation policy was primarily internally driven, particularly by the perceived need for ongoing pragmatic development and improvement of regulatory policy. At the same time, budgetary concerns, positive experiences from evaluations carried out previously in individual agencies and institutions (bottom-up learning), as well as political pressure also play an important role in facilitating regulatory evaluations in several countries.



Source: OECD (2003): Survey of *ex post* evaluation of regulatory tools and institutions in OECD Member Country

33. Governments' increasing pursuit of *ex post* evaluation of regulatory policies confirm the general trend toward more comprehensive evaluations. The survey carried out as part of this report showed that in most OECD countries, rather than being stand-alone policy documents, governments' strategies to the evaluation of regulatory tools and institutions are part of broader regulatory or evaluation policies.

34. The approaches to the evaluation of regulatory tools and institutions seem to mirror countries' overall approach to regulatory management. Member countries with traditions of elaborate, centrally-defined guidance and formal policy requirements seem to take a similar approach to the evaluation of regulatory tools and institutions. In the same vein, countries where the regulatory framework is developed *ad hoc* and pragmatically, and with an emphasis on decentralised responsibilities and self-assessment seem to have a less formalised approach.

Box 2. Examples of Policies and Strategies to Evaluate Regulatory Tools and Institutions

In **New Zealand**, evaluation and review of regulatory tools and institutions is an integral part of the overall regulatory policy framework. Policy principles and guidance documents stress the importance of evaluating and reviewing the effectiveness of regulation, including regulatory tools and institutions. For instance, New Zealand's Code of Good Regulatory Practice (CGRP) includes requirements to review regulations systematically to ensure they continue to meet their intended objectives efficiently and effectively. The CGRP also calls on regulators to ensure that regulatory measures are designed so that they can be adjusted and updated as circumstance change. In addition, New Zealand's Generic Policy Development Process (GPDP) emphasises the need for a review strategy that considers the types of information needed to assess the impact of the policy and whether the strategy needs to be in place at the time of implementation. The GPDP's checklist for the implementation and review stage of the policy process prescribes that monitoring and evaluation procedures are in place that consider whether i) the regulation is effective in achieving the policy objective; ii) the regulation is efficient; and iii) there are any unintended effects.

In the mid 90s, **Canada** introduced quality assurance to its regulatory process to enhance compliance with the Regulatory Policy, through a standards-based approach for assuring that sound systems were in place for the regulation-making process. Canada introduced the Regulatory Process Management Standards (RPMS) but also took on providing guidance for implementation and assessing government-wide capacity. Standards were a way to increase accountability for compliance in departments and to do so earlier in the process (not rely solely on an end-of-the-pipeline check). In 2001, departments and agencies undertook an initiative to develop a common understanding of performance measurement in the context of the Regulatory Policy, to take stock of current "regulatory performance measurement" practices in the management of regulatory programmes and to identify potential indicators. While there is no explicit, government-wide strategy in place which focuses on regulatory programmes (to the exclusion of other programmes), early work is under way to provide institutions with guidance for the application of performance measurement strategies. Ultimately, it is expected that this guidance will provide the basis for a more systematic basis of evaluation data, including information needed for future evaluation of the effectiveness of regulatory tools and institutions.

In **Sweden**, strategies to evaluate regulatory tools and institutions are not expressed in one particular policy document. Several provisions in laws and subordinate regulations include requirements for *ex post* evaluation. Following the Swedish administrative tradition with a high degree of decentralised responsibility to each department/ministry and agency each department and agency has a responsibility to continuously evaluate the regulatory tools and institutions it applies. Evaluations are not centrally coordinated. The intensity of efforts depends very much on the political interest.

2.2. Form of evaluations

35. Evaluation literature identifies three basic types of evaluation (McNamara, 1998). These are process based, goals based and outcomes based. Each type serves a distinct purpose and makes a distinct contribution to the understanding of performance. In broad terms:

- *Process based evaluation* is based on obtaining a clear understanding of the programme's dynamics – that is, it attempts to determine how particular outcomes are achieved. These types of evaluation are likely to be useful where a programme's substantial complaints are experienced regarding the programme or where obvious and significant inefficiencies in relation to programme delivery exist. A process based evaluation is likely to support attempts to identify and address specific shortcomings in the programme through the improved design of specific elements or the addition of other elements.

- *Goals based evaluation* is based on measuring the performance of a programme against the specific goals identified for it at the time of its inception. It answers the question "did the policy deliver what it promised?

– *Outcome based evaluation* measures programme performance against a broader set of measures – looking at whether the programme as a whole has delivered on the underlying objectives of the agency, rather than the specific (and sometimes intermediate) goals that may have been specified for the particular programme.

36. Evaluations of processes are most easily applied when they simply check compliance with formal standards in a "checklist" approach: this approach is inexpensive and methodologically not very demanding; evaluations of goals and outcomes are more complicated and, depending on the object being evaluated, require a certain level of information availability and data processing; outcome targets and measures are becoming more popular. The trade-off is that they cannot be so readily tied to the responsibilities of an organisation or an individual as output measures can.

37. How is this mirrored in governments' efforts to evaluate regulator tools and institutions? Firstly, the distinction above between process, output and outcome focus is congruent with the categorisation of governments efforts to *ex post* evaluate regulatory tools and institutions, cf. next chapter. However as will be shown in the following chapters, there is no strong trend towards preferring one particular approach to the other.

38. The survey of OECD Countries' practices with *ex post* evaluation of regulatory tools and institutions shows that the specific subjects and scope of *ex post* evaluations of regulatory tools and institutions vary. The most commonly evaluated tools are RIA, consultations procedures and simplification mechanisms. The most frequently evaluated institutions are independent regulators and enforcement agencies. Some evaluations are broad based, embracing different sectors and regulatory areas, while other evaluations focus on the application of regulatory tools or institution in one specific area.

2.3. Choosing the appropriate evaluator

39. The organisation of evaluation activities reflects the role governments perceive for evaluation. The question of who evaluates is important in terms of the effectiveness of evaluation (in the specific sense of its ability to analyse programmes or policies adequately), the legitimacy of evaluation and its practical impact – that is, the extent to which action is taken in accordance with the findings of the evaluations conducted. Evaluations can be carried out by a range of actors:

- They can be conducted internally – i.e. by the body responsible for the particular programme or policy;
- by a quasi-independent actor in the administration, such as the regulatory reform body or the Auditor-General;
- By another branch of government, such as a parliamentary committee; or
- By an external consultant, engaged by any of the above parties.

40. The choice of actors clearly affects the likely benefits and drawbacks of the evaluation conducted. Internal evaluation will probably identify and address relevant issues and perhaps tend to produce the most practical recommendations, since the evaluators will have a close familiarity with the issue and good access to programme staff. However, the apparent risk is that there will be a reluctance to reach critical conclusions where programmes are failing substantially and there will be a very limited probability that recommendations to terminate a programme will be made. The legitimacy and validity of internal evaluations may also relatively easily come under doubt.

41. By contrast, independent audit offices or regulatory reform bodies can address more fundamental issues in relation to outcomes, but their choice of issues, findings and recommendations may not necessarily be politically relevant, be it politically or for the improvement of the programme. Similarly, limited access and expertise in specific programme issues may mean that important issues are not reliably identified. Review by parliamentary committees and the like are likely to share the same characteristics.

42. The use of external consultants may have significant benefits in defined circumstances. For example, where consultants with specific expertise in the programme area are engaged by regulatory reform bodies or auditors, the result may be a strong combination of specific programme or policy-related expertise and an independent perspective. Another possible approach may be for evaluations in some cases to be managed by "steering committees" bringing together both programme managers and regulatory reform bodies or other representatives with an independent perspective.

43. Seen from a general public policy evaluation point of view, the pros and cons of different approaches in illustrated below.

Table 1. Who Evaulates? Advantages and Disadvantages of Different Approaches

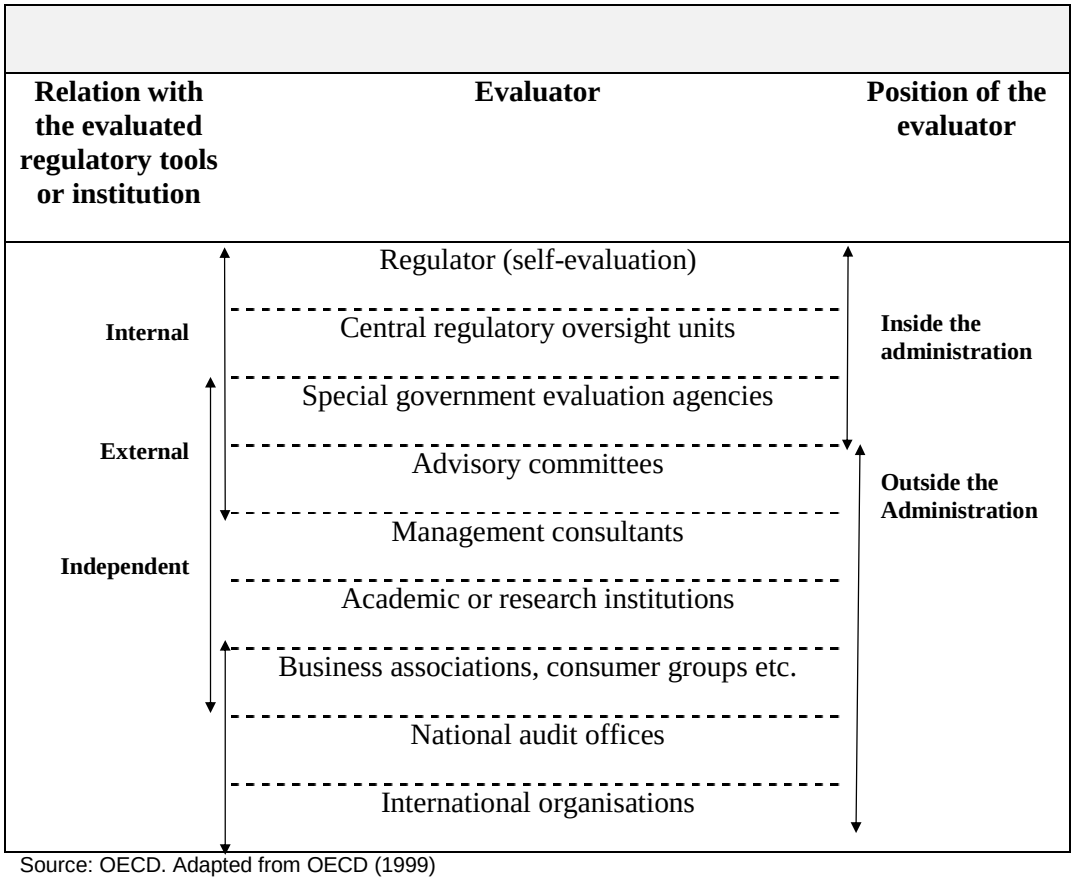
	Self-evaluation	**Independent evaluation**	**Participatory evaluation**
Advantages	Maximises learning	Competence Legitimacy Speed	Lessons applied
Disadvantages	Can avoid difficult issues	Limited impact Low dissemination	Low competence Requires commitment Slow

Source: Forss K. (2004)

44. Answers to the background survey of OECD country practices on *ex post* evaluation of regulatory tools and institutions show no clear trend to favour the use of any particular evaluators. Different countries use different solutions. In some countries, specific units or agencies are primarily responsible for monitoring and implementing evaluations of regulatory tools and institutions. In other countries, the Ministry of Finance and/or national auditor institutions have the primary responsibility for evaluations. There are also examples where the evaluation policy is being co-ordinated from the central regulatory oversight unit located in the prime minister's office. Finally, in several countries, the responsibility for conducting evaluation lies within individual ministries. It is notable that, where responsibility for evaluation lies with the ministry in which the programme or policy being evaluated is found, the evaluation itself is frequently outsourced. By contrast, where independent audit agencies are responsible for evaluation functions, they often also carry out the evaluations.

45. Seen from a regulatory point of view, it is possible to nuance to the list of potential evaluators and their association to the evaluated regulatory tool or institution, cf. Table 2 below. The box also gives an indication of which of these can be seen as internal, external or independent. Depending on the perspective there is an overlap between these categories.

Table 2. Positions of Evaluators and Their Relation with the Evaluated Regulatory Tool or Institution



Source: OECD. Adapted from OECD (1999)

2.4. Methodological issues

46. There is no best practice in terms or evaluative process or method. Different approaches and methodologies have different advantages and drawbacks and the relative importance of these varies with different programme types and contexts and with different decision-making processes. Various ways of collecting and analysing data provide different perspectives on the evaluated programme. There is, for example, a considerable discussion on the advantages and disadvantages of using quantitative and qualitative methods in evaluation. The choice of methodological approach also depends on various factors such as the available resources, access to data and technical expertise.

47. Methodological problems are intrinsic in all approaches to evaluation. For instance, problems related to causality are common to social sciences in general. Conclusive evidence of cause-effect relationships can rarely be established, since controlling all relevant variables is seldom possible. Choosing criteria for evaluation may be problematic simply because the intended objectives of public programmes are often multiple, vague, hidden, evolving and even conflicting.

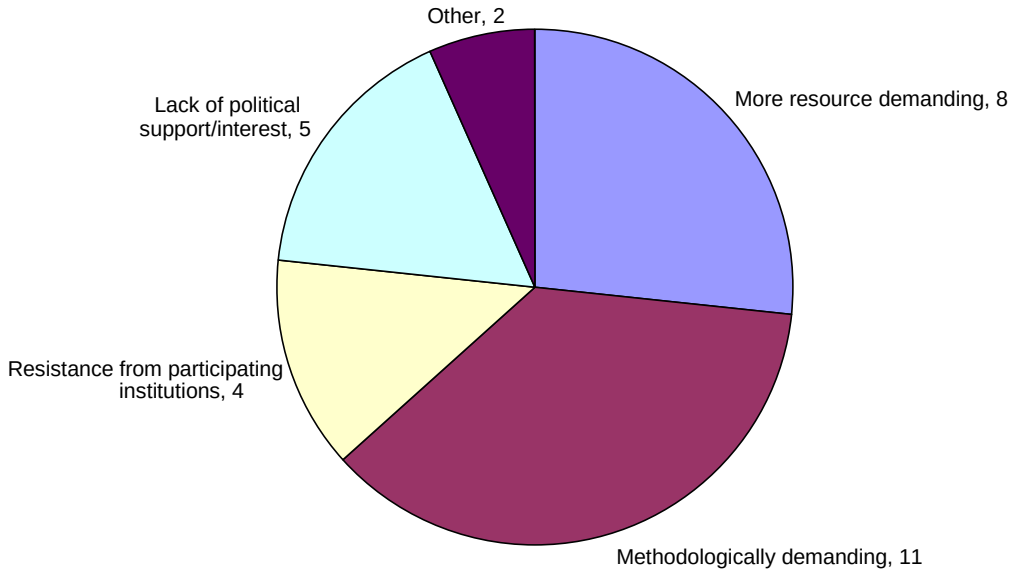
48. Methodological problems in evaluation can be dealt with when limitations are recognised and problems properly addressed. This requires specific knowledge and skills that can be gained by training staff and commissioning external expertise to conduct evaluations. Also, combining different methods is often the most fruitful approach to lessening methodological problems in evaluation. Furthermore, appropriate quality control mechanisms (see later discussion on ensuring technical quality) can be set up to guide the evaluation process.

49. What are the practices and experiences that can be observed in terms of governments' effort to evaluate regulatory tools and institutions? Surveys and case studies seem to be the preferred methodological approach. Case studies, with an extensive examination of a small number of cases, have the advantage of being able to give attention to case-specific details and exploratory use of data. Generated hypotheses cannot be tested in a scientifically rigorous way because the number of cases is limited and chosen non-randomly. Statistical analysis allows drawing statistical inferences on the basis of a large number of cases. However the applicability of statistical analysis to regulatory tools and institutions is rare, since the precondition of a reasonable sample is fairly uncommon. Quantitative analysis with descriptive statistics or regression analysis of regulatory institutions (regardless of their organisational specifics or decision-making structure) is feasible, but methods require a certain level of data availability and are thus not frequently used.

50. Another methodological characteristic of governments' *ex post* evaluations of regulatory tools and institutions is that they are often not carried out on a systematic basis. Seemingly – and perhaps for that reason – there is little learning over time and across countries and policy areas.

51. OECD countries seem to face similar challenges in their pursuit of evaluating regulatory tools and institutions. Most notably, countries point to methodological difficulties as some of the most significant and unforeseen problems they have encountered. Resource constraints, lack of political support, and resistance from participating institutions also cause difficulties. At the same time, the approaches taken in many evaluation projects are developed *ad hoc* with the generation of no or very little quantitative or comparable data. Thus, there seems to be significant scope for the dissemination and further development of evaluation methodologies.

Figure 2. **What have been the most significant challenges encountered in the process of assessing regulatory tools and institutions?**
(multiple answers possible



Source: OECD (2003): Survey of *ex post* evaluation of regulatory tools and institutions in OECD Member Country

52. The evaluation of regulatory policies is also sometimes constrained by the mixed incentive among those (potentially) carrying out the evaluations. Compared with a business organisation operating in a competitive market, institutions entrusted with the development, implementation and enforcement of regulatory policies tend to lack strong performance incentives. Objectives tend to be diffuse (cf. the more focused profit objective of business organisations) and there is no 'outside' mechanism to evaluate the likely future effects of decisions taken now (cf. capital markets, which can quickly incorporate the future, anticipated consequences today's decisions in asset valuations). Linkages between decisions, effects/consequences and objectives are therefore difficult to determine, creating problems of accountability and oversight, even where there exist well-established institutional mechanisms dedicated to these latter purposes. (Systematic) evaluations of regulatory policies may make such incentives stronger.

53. Regulatory decisions involve bureaucratic budgets, private incomes, and political power. Reliable data, or even the availability of data, are not guaranteed in such situations. Incentives to avoid recording data involve the interests of governments to “speak with one voice,” objectives may be conflicting, there can be interests in obscuring the political trade-offs that occur, and some data can be relatively non-standard such as tracking text changes in documents or the nature of meetings that occur.

54. The “causality issue” is also at play when it comes to the evaluation of regulatory policies. Regulatory processes have multiple players such that the simultaneous activity of the parties may make identification of individual impacts difficult. For instance, an agency may interact with the central regulatory authority in many ways so that it can be difficult to separate the product of the originating agency and the review agency. In such circumstances, only the combined final product may be observable.

3. A CONCEPTUAL FRAMEWORK FOR *EX POST* EVALUATION OF REGULATORY TOOLS AND INSTITUTIONS

55. This chapter proposes a conceptual framework within which the *ex post* evaluation of regulatory tools and institutions can be considered. The framework has been developed and adopted in response to a number of specific needs in relation to evaluation of regulatory tools and institutions:

- First, the research undertaken to date, including responses to the survey of member countries, reveals that existing country practices vary widely and are not immediately comparable. Thus, a conceptual framework can be considered as a means of organising our knowledge of existing practices, bringing a consistent analytical perspective to bear and helping enhance the comparability of the different approaches;

- Second, the adoption of a conceptual framework will assist in clarifying and making explicit the links between the various evaluation tools and mechanisms and the underlying objectives of the evaluation process. This, in turn, assists the development of a critical analysis of existing practice and the formulation of proposals for further development;

- Third, the evaluation process is itself likely to be subject to dynamic evolution. As experience with the use of regulatory quality tools and institutions accumulates, so the range of objectives evaluation can reasonably attempt to pursue will expand. Application of the conceptual framework will assist in clarifying the evolution of evaluation practice over time.

56. The conceptual framework is in three parts, although there are inevitable links and overlaps between these basic elements. The three types of evaluation are as follows:

- *Compliance tests* seek to evaluate formal compliance with the individual elements of the regulatory quality tool or institution in question. That is, they test whether the RIA process, the consultation process, or the regulatory institution in question has meet the procedural requirements set out in laws, policies or guidelines as appropriate. These tests are essentially process focused. Compliance tests can be applied on both an ex ante and an ex post basis.

- *Performance tests* essentially measure the quality of the analysis undertaken, going beyond the question of formal compliance with procedural requirements. Performance tests seek to evaluate the performance of the regulatory tool or institution in terms of its ability to add sophistication and relevant data to the regulatory development process and so support a high quality regulatory process. Some performance tests can be carried out *ex ante* – such as assessing the assumptions of an economic analysis. However most performance can only be carried out *ex post*, since it involves assessing the accuracy of the ex ante analysis by comparing it with the results that were actually observed in practice. For example, did RIA correctly predict actual benefits and costs? Did consultation identify alternative policy options, or gather substantial new data on which to base decision-making? Performance tests can be considered as "output focused".

- *Function tests* seek to evaluate the actual effect of the regulatory tool or institution on the quality of the regulatory outcome. Thus, they are effectively measuring not only the regulatory

quality tool itself, but also the degree to which it is effectively integrated into the policy process via functioning "policy feedback" loops. These tests can therefore be considered as "outcome focused", and can only be carried out *ex post*.

57. Each of these tests is intended to evaluate the regulatory tool or institution according to a different, and progressively more demanding criterion. Compliance tests essentially measure whether the tools are taken seriously by regulators. Their concern is with the inputs to the regulatory quality requirement. Performance tests link the inputs with the outputs and ask whether the regulatory quality tool is functioning adequately. Function tests also include the link between the quality tool and the policy process itself and focus on whether the tool is being used *effectively* in the sense of contributing to improved policy outcomes.

3.1. Compliance tests (process)

58. Compliance tests are evaluations checking that regulatory tools and the workings of institutions are applied in accordance with formal standards and requirements. For example, such assessments ask whether RIA or consultation mechanisms meet the applicable guidelines, or, in the case of regulatory quality assurance units, whether regulators have tabled draft regulations for reviews by this unit in accordance with internal government procedures.

59. This "checklist" or "scoring" approach applied in compliance testing has the advantage of providing a transparent, easily comparable and efficient check-up on the performance of regulatory tools and institutions. It is the most readily available performance indicator for measuring "administrative success". The value of such tests is based on the obvious presumption that compliance with the government's regulatory quality guidelines is essential if they are to play their intended part in supporting regulatory quality. Clearly, if regulatory agencies are not complying with even formal requirements to adopt regulatory quality tools, fundamental action is required to address this failure of implementation at the "first hurdle".

60. As well as the advantages of comparability and transparency, compliance tests are probably the least expensive and methodologically demanding of evaluations. Despite this, they are clearly of value, since many of those that have been conducted in a range of countries have highlighted high levels of non-compliance with regulatory quality requirements particularly, though not exclusively, during the early stages of their implementation. Compliance tests can therefore be useful both in focusing efforts by regulatory reform authorities to enhance compliance levels and in measuring improvements (or deterioration) over time.

61. However, while presumably a necessary condition, compliance with formal regulatory quality requirements is not sufficient to ensure that there is a real positive impact on regulatory quality outcomes. It is a commonplace observation among regulatory reformers that regulators are capable of achieving a high level of formal compliance with quality tools without their impinging significantly on the nature of the regulatory process or its final results. Formal requirements for the application of regulatory quality tools tend to be relatively general and "high level" in form. Thus, compliance tests tend to measure whether a particular required action has been carried out, rather than whether it has been carried out with an acceptable level of quality. The latter is the focus of performance tests.

3.2. Performance tests (output)

62. Performance tests move beyond the threshold question of whether the compliance tests are actually applied to focus on the quality of that application. Many of the criteria for these assessments will be found in the supporting "guidance" material on compliance published by regulatory reform authorities

and other central agencies. Such guidance tends to be framed in "advisory" terms, rather than as mandatory requirements, often reflecting that elements of judgement are required in determining how to apply the regulatory quality tool to particular circumstances. Nonetheless, a global view of these guidance documents can provide a sound basis for the application of performance tests. For example, RIA guidance may set out several acceptable costing methodologies. A compliance test might judge a RIA as meeting quality standards if any of these is adopted, but a performance test may regard their incorrect adoption, or the use of another, unacceptable methodology, as representing a failure to meet the standards.

63. Some performance tests can be carried out both *ex ante* and *ex post*. Test carried out *ex ante* can focus on whether methodological elements are correctly (or at least defensibly) applied in the given circumstance: is a plausible discount rate adopted and is it applied consistently? Does the treatment of costs and benefits avoid double counting errors? Are costs adjusted to remove the effects of taxes, subsidies and like distortions? Does consultation documentation provide appropriate information and data to interested parties? Are response times adequate? Has there been substantial feedback from stakeholders? Thus, these tests seek to ensure that the formal requirements have been acquitted competently and professionally. Obviously, such tests can also be carried *ex post*, often benefiting from the availability of data from a number regulations already implemented, as opposed to *ex ante* tests, which would normally focus on tools applied in the context of one individual regulation.

64. Ex post performance testing is focused on the predictive ability of the quality tool applied. That is, were the conclusions of the RIA, or the predictions received in the consultation process, borne out by practical experience with the implementation of the regulation? At a basic level, this form of testing appears conceptually fairly simple and unambiguous: to the extent that clear predictions of regulatory impacts are made *ex ante*, the task is simply to compare them with observed reality. However, closer examination suggests that complications abound. The *ex post* performance data is rarely (readily) available. Moreover, it is often unclear to what extent complicating environmental factors not able to be predicted by RIA authors of consulted parties intervened to affect outcomes. While summary analyses that seek to compare predictions and outcomes across a wide range of regulations may "smooth out" these effects and help reveal systemic biases, it remains possible that different results will be observed in different time periods, due to more widespread confounding factors.

65. An important potential benefit of this kind of ex post testing is the possibility that it will lead to the identification of systemic factors that are reducing predictability and, in turn, allow these to be mitigated by actions to improve guidance material provided to regulators, improve the scrutiny exercised by regulatory reform bodies or, at a minimum, apply "adjustment factors" to ex ante predictions made. In more general terms, the use of ex post testing provides an operationalised performance test to the regulatory quality tool or institution in question. That is, a high "predictability score" reinforces the value of the tool within the policy context as an aid to rationally based decision-making. Similarly, a low score is a strong indicator of a regulatory quality assurance system that is not meeting expectations and is likely to require attention to the diagnosis of its weak-points and opportunities to address them.

66. Another critical performance dimension which could be subject to evaluative tests is *timing.* [3] Many case studies have documented the critical importance of timing to the usefulness of regulatory tools.

[3] Tests focussing on the timing of the application of the regulatory tool could to some extent also be labelled as compliance tests, since timing is clearly a function of the design and implementation of the regulatory quality tool, or as function tests, since timing is also a symptom of a tool's (non-)integration with the policy process in practice. However in the context of this report timing tests are characterised as a performance test. Partially because government guidelines rarely specify exact criteria for when to initiate the application of a regulatory tool (they normally say "as early as possible"), which could be subject to the checklist approach applied under compliance tests. Partially because timing *per se* does not have the same indicative value of the outcome of the regulatory process, as have other tests characterised as function tests.

If, for example, RIAs and consultations are not initiated until after the regulatory process is well underway, these tools obviously have difficulty being influential. Not only is the "bad news" of a poor benefit/cost outcome delivered after a high degree of commitment to a particular regulatory approach has been made, but pressure on the analyst not to deliver such bad news about benefits and costs may mean that these tools are not applied with appropriate rigour and impartiality in the first place, thus creating cynicism about the nature of the process and the value of the tools.

67. However, a strong result in terms of ex post performance testing does not necessarily lead to the conclusion that regulatory quality tools and institutions are having real, substantial and positive effects on regulatory quality outcomes. Lack of "outcome" performance can be the result of a failure to integrate high quality regulatory tools and institutions with the policy decision-making process. Identifying this problem of linkage, and providing the tools to analyse it, is the role of *function*, or outcome tests, considered next.

3.3. Function tests (outcome)

68. Function tests are concerned with the effect of regulatory quality tools on regulatory outcomes, that is, whether the application of these tools and institutions contributes to improving the quality and performance of regulations in practice and the quality of the regulatory decision-making process. Test of this sort are fundamental to answering the question of whether regulatory quality tools and institutions are delivering on their promise of improving the effectiveness, efficiency and accountability of government regulatory actions. If performance on function tests is poor, the benefit of the – usually very substantial – investment made in implementing the regulatory quality tools is inevitably minimal and substantial diagnostic efforts are required to determine the source of the failures and suggest the means of addressing them.

69. A range of function tests can be identified. An approach that is widely applicable is that of the *"audit trail"*. Here the test attempts to determine the responsiveness of the tool or institution to the inputs received. Thus, in the case of RIA, tests would seek to verify if new options identified as part of the process have been properly and consistently evaluated against initially identified options and whether, in the case of apparently superior alternatives being identified, changes to the policy proposal have resulted. In the consultation context, such a test would similarly question whether proposals raised were given due analysis and, where merited, integrated into policy development. A similar test can be proposed in the case of regulatory reform institutions, to determine whether their inputs to the regulatory process (e.g. review comment on draft RIA) have been influential in terms of final outcomes.

70. The *incidence of actual policy changes* in response to information obtained via RIA or consultation is a relatively easily measurable form of function test. It is reasonable to expect that regulatory organisations should be able to provide relatively detailed examples of ways in which third party inputs into RIA or consultation processes have materially influenced decision making. Explicitly recognising examples of responsiveness in a positive light may also have the desirable property that it acts to counter any potential tendency in regulatory agencies to treat such responsiveness as a signal or admission of prior failings. However incidences of policy changes can be misleading for several reasons. First, a low incidence of change can suggest a non-responsive policy apparatus. However, it can also suggest one which is functioning at a high level, such that initial policy proposals are well researched and articulated and infrequently require major change at later stages of development. Only a deeper review of the broader question of responsiveness is likely to show conclusively which effect is predominant. Second, measures of visible policy changes are likely to under-estimate the effectiveness of policy disciplines such as RIA, consultation and central oversight bodies. This is so to the extent that the presence of these disciplines itself imposes changes on policy-making behaviour within line agencies. That, is more rigorous testing of

proposals may occur within departments and less robust proposals may be less likely to go forward to the quality assurance stages due to a desire to avoid subjecting them to criticism that they may not survive.

71. A broader approach to function testing would be to attempt to measure *changes in the administrative culture* among regulators as a consequence of working with various regulatory tools and institutions. This is necessarily a more judgemental and qualitative approach. However, given the expressed intention of these regulatory quality initiative to promote such long-run change, these sorts of tests can be seen as central to the measurement of their success. Examples of such broader reviews – or at least of the question of cultural change being addressed as a part of the process of reviewing regulatory quality tools – have been reported by a number of Member countries, as will be described in the following chapters.

3.4. Conclusion

72. A clear conclusion of the above discussion is that the different approaches to evaluating regulatory tools and institutions are not substitutes but complements. While only function tests are able to answer the question of whether evaluation tools are affecting achieved regulatory quality, compliance and performance tests allow diagnosis of particular aspects of the implementation process, while also in many cases being a less resource-intensive means of obtaining useful information about the implementation and performance of regulatory quality mechanisms than are function tests. Thus, compliance and performance tests can be used as a basis for "fine tuning" the implementation of regulatory quality mechanisms by highlighting key points of departure from initial designs or departures from international good practices.

73. Previous OECD work has suggested that the implementation of regulatory quality tools is a long-term process and that a progressive broadening and deepening of the requirements of tools such as RIA constitutes an appropriate strategic approach to this long-run implementation issue. It seems likely that a similar view can and should be taken in relation to the various evaluation tests discussed in this section. For example, in the early stages of RIA implementation, a key focus must be on achieving a high level of formal compliance with the requirements. Thus, compliance testing should be a central evaluative concern. By contrast, it is unlikely that a newly implemented RIA process will have had substantial opportunities to affect policy outcomes and the administrative culture among regulators. Hence, the adoption of function testing may be a poor use of resources at this stage. On the other hand, a stronger focus on performance and functional tests is clearly necessary in the case of mature RIS or consultation processes.

74. It follows from this view that the above taxonomy of evaluative tests can be used by regulatory reform authorities as the basis for organising evaluations and timing the application of the different tests and approaches. As well, this taxonomy can be seen as forming a kind of "checklist". Given the view that the three basic types of test (compliance, performance, function) are complementary in nature, regulatory reform authorities can potentially approach the consideration of evaluation efforts from the viewpoint of ensuring that each kind of test has been implemented by the time the regulatory quality tool being evaluated is itself in a "mature" implementation phase.

75. Notwithstanding the above, evaluation efforts must ultimately be assessed in benefit/cost terms, consistent with the general objective of regulatory governance of maximising overall (i.e. societal) welfare. Thus, assessments should remain focussed on the overall target or determining whether and how regulatory tools and institutions are helping to optimise policy in practice. That is, do they assist in ensuring that the benefits from regulation or other policy initiatives are maximised, and the costs minimised. The benefits following from evaluations must outweigh the costs, where benefits are counted in terms of an effective "feedback loop", whereby the functionality of the regulatory quality tool or institution is enhanced as a result of the evaluative process being undertaken. However, in applying such a criterion, it is essential to

recognise the need to go beyond static considerations. The benefits accruing from evaluation may incorporate important dynamic and cultural effects, arising particularly from the spread of learning regarding the effectiveness of different tools across a range of policy areas and a potential responsiveness to this learning among regulators at the regulatory design stage.

76. While benefits due to evaluation can be difficult to account for fully, difficulties in relation to the costs of such activity also arise. On the one hand, it is clear that substantial costs, time and resources are likely to be required to produce much of the data necessary to completion of some types of evaluations. On the other, it is often the case that, once a framework for collecting and reporting data has been established, incremental costs may be marginal. This once again points to the importance of institutionalising regulatory policies, including ex post evaluation, thereby standardising data and reporting requirements and ensuring that the basis for high quality evaluation is laid at the time that regulation is implemented.

4. REGULATORY IMPACT ANALYSIS

4.1. Introduction

77. This section reviews the experience of Member countries in evaluating Regulatory Impact Analysis (RIA) and integrates this experience with some theoretical work on possible evaluative objectives, approaches and outputs. It presents this material within the context of the analytical framework developed in Section 3; that is, distinguishing compliance, performance and function-based evaluative tests.

78. RIA is central to the set of regulatory quality tools used by OECD governments. Its use is closely integrated with other regulatory quality tools, notably with public consultation and the role of central regulatory reform oversight bodies. Indeed, it arguably constitutes an amended rule-making process, which contains all the elements of rational public action (Harrington 2004). In some Member countries, there is substantial experience in the implementation of RIA, stretching back to the late 1970s and early 1980s in a few cases. In many more Member countries RIA was implemented during the latter half of the 1990s as the use of this tools spread rapidly to reach the current situation in which virtually all OECD governments use some form of RIA. Given this extensive base of experience with the practical implementation of RIA, there is a clear need for evaluative tests of all of the above kinds to be employed if a full evaluation of the acquired experience is to be made.

79. RIA is generally seen as a tool which favours rationally based decision-making over other forms such as political, expert driven, or consensus based approaches (see OECD 2003). This again suggests that a high level of evaluation activity would be expected, since a rigorous and rational view of the use of such a tool would be consistent with the nature of the tool itself. In practice, however, the use of evaluative tests in relation to RIA is less common than might be expected. Harrington (2004, p1) concludes that *"…there have been relatively few systematic attempts to evaluated the RIA requirements themselves to determine whether RIAs as a group have actually lived up to their billing"*.

80. Possible reasons for these observations include the fact that the process of grafting RIA on to, and integrating it into existing governance systems means that, in practice, the tool is not always established – or used – as a tool of rational decision-making as theoretical observation of the nature of the tool would suggest. The experience of the OECD country reviews suggests that the institutional and bureaucratic contexts for RIA vary significantly in the context of different governance traditions. Thus, where the Anglophone countries have generally emphasised a strong link between the use of RIA and a quantitative, strictly "rational" approach to decision-making, practices in Europe are more likely to accommodate the fact that different stakeholders bring different perspectives and indeed different logics to the RIA process. Countries with corporatist patterns, such as Denmark and Germany, have to some extent re-interpreted RIA as another instrument of consensus-generating negotiation. This is why in some countries RIA does not produce a final set of figures that purport to show conclusively whether the benefits justify the cost of the proposed regulation, but rather a set of partial estimates that are then used by policy-makers in a mode that is more negotiation than technical analysis of options. (See also Radaelli, 2003) The difficulty of adapting the RIA model – pioneered in the Anglophone countries – to these different governance traditions may, in turn, be one reason for the relatively slow rate of adoption of RIA within many European countries.

4.2. Compliance tests

81. As set out in Chapter three, compliance tests are evaluations checking that regulatory tools and institutions are applied in accordance with formal standards and requirements. In the case of RIA, such assessments typically ask whether the RIA meets the applicable guidelines for preparation of RIAs.

82. While these guidelines vary to some degree between countries – particularly reflecting differences in terms of whether the RIA model adopted requires full benefit/cost analysis or a partial approach, as well as the degree of integration of RIA and public consultation – there is a high degree of observed commonality, cf. Box 3.

Box 3. Common Characteristics of RIA

1. *Statement of problem*. Is government intervention both necessary and desirable? Is the problem or condition unlikely to be resolved by the intervention of private?
2. *Definition of alternative remedies*. These include different approaches, such as the use of economic incentives or voluntary approaches.
3. *Determination of physical effects of each alternative, including potential unintended consequences*. The net should be cast wide. Generally speaking, regulations or investments in many areas of public policy can have environmental implications that must be kept in mind.
4. *Estimation of benefits and costs of each alternative*. Benefits should be quantified and where possible monetized. Costs should be true opportunity costs, not simply expenditures.
5. *Assessment of other economic impacts*, including effects on competition, effects on small firms, international trade implications.
6. *Identification of winners and losers*, those in the community who stand to gain and lose from each alternative and if possible, the extent of their gains and losses.
7. *Communication with the interested public*, including the following activities: notification of intent to regulate, request for compliance cost and other data, public disclosure of regulatory proposals and supporting analysis, and consideration of and response to public comments.
8. A clear choice of the preferred alternative, plus a statement defending that choice.
9. *Provision of a plan for ex post analysis of regulatory outcomes*. It is important, for example, to establish current conditions to have a benchmark to measure performance against. Planning is needed to ensure that procedures are in place for the collection of data to permit.

83. This kind of detailed set of guidelines appears to provide a relatively simple "benchmark" for use in compliance testing. Arguably, one could assume that RIA compliance tests would be the most frequently applied *ex post* evaluations of regulatory tools, given the immediate availability of input data (the RIA documents) and the rather simple evaluation criteria (were process x, y and z carried out in accordance with the relevant requirements?). However the OECD survey carried out as part of this report suggests that RIA compliance tests are not among the most frequently applied tests that governments use to evaluate RIA performance. It is notable, that most RIA compliance tests are carried out by non-governmental organizations.

84. In a majority of cases, these compliance tests indicate widespread non-compliance, notwithstanding the relatively clear critieria that are usually established for RIA compliance. For example,

Hahn has undertaken a series of RIA evaluations in the United States over a number of years, based on this form of benchmarking actual RIA against published guidance material. In a review of 48 RIA completed between 1996 and 1999, Hahn (2000) found that many were not only deficient in items that were considered to be essential to the production of a quality output, but also failed to conform to the relevant RIA guidelines (issued by the Office of Management and Budget) in important respects. Hahn found that while 90 per cent of RIA monetized costs, only 50 per cent monetized benefits and only 29 per cent calculated net benefits.

85. An extensive survey of British RIA (Ambler et al. 2003)[4] provides similar results. Moreover, its comparison of results across different periods (i.e. 2002-03 vs 1998-2002) indicates limited progress in achieving enhanced compliance over time. Costs were reported in 58 percent of cases in 2002-03, compared to 55 in the earlier period. Benefits were reported in only 9 percent of cases (6 in the earlier period). The authors report that often the "benefits" section consists of little more than statements of the form, "Consumers will benefit from improved safety standards," with no effort at quantification. While most RIAs do contain a report on the consultation process, it is often couched in generalities and does not say how the regulation was altered in response to the comments. The one place where the report found improvement over the earlier period was in the consideration of alternative regulatory approaches. The proportion of RIAs mentioning non-regulatory alternatives doubled, from 11 to 23 percent. Overall, the authors conclude that "The impression remains that in many cases completion of RIAs remains a bureaucratic task to be despatched with as little effort as possible."

86. Another example of a RIA compliance test carried out by a non-governmental organisation is the study carried out by the Swedish Board of Industry and Commerce for Better Regulation (NNR). Since 2001, the NNR has published an annual report on the compliance of Swedish government RIA with 11 "quality factors". Most of these quality criteria are identical to the requirements laid down in government guidelines, while additional criteria reflect aspects of RIA quality that the NNR considers significant but which are not currently included in the guidelines[5]. Although the 2003 study identifies progress compared to 2002, these improvements come from a very low base. For example, only half of the RIA comply with the statutory requirement to draw up a special RIA for small and medium-sized enterprises (SMEs). In 2003 37% of all RIAs included descriptions of possible alternatives (up from 26% in 2002). The number of RIAs reporting the expected total regulatory costs were five per cent in 2003, up from 4 per cent in 2002.[6] (NNR, 2003).

87. How can the apparently low level of compliance with basic RIA requirements be explained? In the British and Mexican cases, it can be noted that the RIA requirement is a relatively recently established one and that compliance can be expected to take time to establish – even at the basic level of formal compliance with guideline requirements. Clearly, however, the situation is different in the United States, which was the earliest adopter of RIA requirements. One suggestion made by Harrington (2003), in

[4] 165 RIA publishced between July 2002 and June 2003 were assessed, of a total of 197 published during this period.

[5] The inclusion of these additional criteria can itself be seen as an indicator that the government's RIA guidelines may be in adequate in several respects and meriting further scrutiny. That said, it is perhaps unsurprising if a lower level of compliance were to be found in respect of these additional criteria, given their non-inclusion in the guidelines.

[6] The 11 quality factors used in NNR's study are: 1) Summary of proposal, 2) Reference/summary of previous regulation in the area (if any), 3) Alternatives described, 4) Reporting on the manner in which consultations with effected companies and sector organisations have taken place, 5) Reporting on the number of companies effected by the proposal, 6) Reporting on the total costs imposed on individual companies, 7) Reporting on total costs, 8) Reporting on competition aspects and effects of proposal, 9) Reporting on relation of proposal to the EU, 10) Circulation of proposal for comments for at least three weeks, and 11) RIA carried out in accordance with the *Simplex Ordinance* (required if proposal has impacts on SMEs.

reporting the Hahn results, is that it is possible that the more rigorous, higher quality RIA are correlated with the more important and far reaching regulatory proposals. This is intuitively plausible since regulators can be expected to devote a larger part of their RIA resources to such tasks – an approach encouraged in a number of OECD countries, such as the Netherlands, which have introduced a two-step RIA procedure, allowing for more intensive tests when regulatory impacts are considered significant. More broadly, several guideline documents specifically counsel a focus of resources on farther-reaching regulatory proposals, on the basis that it is likely to represent the most effective use of resources. It can also be argued that resources should be targeted where the impact of the proposed regulations is largest and where the prospect of altering regulatory decisions was greatest. Thus, it is plausible that, while there is a relatively low level of compliance with guidelines overall, there is a higher level of compliance with regard to major RIA.

88. A second issue is that of the feasibility of quantification and monetisation of costs and, in particular, benefits in relation to many regulatory proposals. The difficulties involved in completing these tasks are widely acknowledged.[7] A sophisticated form of compliance testing might include judgements as to whether, in cases where quantification and monetisation were lacking, substantial further progress was considered feasible. That is, the benchmarking would be undertaken against a specific feasibility criterion, rather than a general requirement. However, such an approach would clearly introduce a larger element of subjectivity into the analysis.

89. A third reason for low compliance with RIA requirements may derive from the sometimes strong political pressures on reform bodies to avoid stalling the regulatory process by insisting on full compliance with RIA processes – particularly where regulatory agencies are unresponsive and may have delayed the commencement of RIA activity until relatively late in the regulatory process. More broadly, the commonly mounted argument that RIA somehow diminishes the legitimate role of Ministers in taking responsibility – and having authority for – regulatory decisions can be seen as providing another pressure on regulatory reform authorities to adopt a less vigilant approach to ensuring compliance.

90. A key element in the use of compliance testing is clearly that of determining the reasons for non-compliance. As noted above, RIA is predicated on the need to achieve a long-term cultural change in the approach taken to regulatory activity. Non-compliance, particularly in the early stages of implementation, may denote a lack of motivation to comply, indicating that the cultural change has not been achieved. The appropriate response may, in part, be to adopt a more rigorous scrutiny and enforcement regime in the short term. Practical limitations on the extent to which compliance is possible may also be important, as noted above. However, non-compliance may also reflect a lack of appropriate skills and expertise, which can be addressed directly by the central regulatory reform institution. Thus, for example, the results of the recent RIA quality review undertaken in Mexico mentioned above were used to assist in identifying areas where regulatory quality is inadequate and therefore aid in targeting training for regulators.

91. It should be noted that compliance testing also can be, and is, undertaken *contemporaneously* with the development of the RIA in a significant number of OECD countries, as this is often an integral part of the RIA process for developing individual regulations. Such review is in some cases undertaken by the central regulatory reform authority, though other bodies that are independent of the regulatory agency also take responsibility in some cases[8]. Formally speaking, approval of the RIA document by these bodies

[7] OECD (1997), for example, argues that such difficulties must be recognised but should not be used as a reason to retreat from the use of BCA, since a partial analysis is, in any case, preferable to the absence of any such quantitative analysis at all.

[8] In some cases, this function can be performed by non-governmental bodies. For example, Victoria (Australia) allows any competent person or organisation to certify that a RIA document complies with the relevant requirements. See: *Subordinate Legislation Act 1994* (www.parliament.vic.gov.au).

is usually a pre-condition for the regulatory process to proceed – a fact which would suggest that a high level of compliance would routinely be observed. However, as illustrated above, the results of retrospective compliance testing suggest that the practical outcome is often different.

92. The inclusion of required RIA components is clearly a necessary but not a sufficient condition for a good RIA. If these items are not present RIA can hardly be considered of good quality, for they speak to matters that are essential for being able to assess regulatory impact and indeed reflect acceptance of the basic parameter of rationally based, comparative policy-making. However, it is equally clear that RIA may not measure up even though they satisfy the checklist. For this reason, performance tests have been developed and applied to RIA. The nature and findings of these tests are reviewed in the following section.

4.3. Performance tests

93. Performance tests are concerned with the quality of the various elements that make up the RIA, and with their internal consistency, rather than simply the question of whether the elements required are actually present. At the most basic level this sort of analysis examines whether the RIA avoids egregious errors, such as double counting of benefits or costs, confusion of costs and expenditures, improper definitions of benefits, failure to distinguish between cost or benefits and transfer payments, improper discounting, etc. It also examines the transparency and clarity of the RIA. Do the authors explain how they arrived at their conclusions? Can quantitative outcomes be linked to inputs? Do the authors make clear what assumptions they are making? Are those assumptions reasonable? Do the authors define an appropriate counterfactual or baseline? This kind of tests can in principle also be carried out *ex ante*, focusing on whether the RIA analysis is considered methodologically adequate and internally consistent at the time of writing. Tests can also focus on timing of RIA, i.e. when in the regulatory process a RIA was initiated. At a more sophisticated level, performance tests focus on the actual regulatory impacts compared to those predicted in the RIA prior to the regulation being introduced. Such performance tests gather evidence as to the predictive abilities of RIA and provide information on the actual performance of regulations in practice. Thus, their role can be seen both as a test of RIA and as a broader tool of evaluation of government policy.

94. Performance tests are the most frequently reported form of *ex post* evaluation of RIA. They are typically carried out as one-off projects, assessing a limited number of selected RIAs as part of, or feeding into, a programme of reform of the existing RIA system. A recent example of performance testing was carried out in the United Kingdom between 2000 and 2002. This work, carried out by the National Audit Office (NAO) and supported by the Cabinet Office's Regulatory Impact Unit, has led to a series of projects aiming at improving the quality of RIAs. In 2000-2002, the NAO examined a sample of 23 RIAs across 13 departments and agencies to study the way in which RIAs were prepared and to identify the scope for learning lessons. The NAO reported examples of good practice by government departments in preparing RIAs. It set out why RIAs are important, the key features of RIAs which add value to policy making and the further steps that departments and the Cabinet Office could take to improve the RIA process. The review was primarily based on interviews with staff responsible for preparing RIAs and staff in regulatory oversight units. As a follow-up to the National Audit Office (NAO), the Cabinet Office reviewed the RIA guidance for policy makers. The review was based, among others, on the results of a survey covering the views of all stakeholders. Box 4, below, contains the questions used to conduct this survey.

Box 4. Reviewing RIA guidelines in the United Kingdom: Survey questions.

1. Are there any specific areas of the RIA where, in your experience, analysis has sometimes or often been insufficient to inform policy makers or consultees?
2. Will the revised guidance direct those writing RIAs to provide a fuller analysis in those areas?
3. If you represent a charity or voluntary organisation, are there any specific areas or issues which should be brought to the attention of policy makers? How can this be best achieved in the guidance?
4. Is the guidance easy to follow?
5. Do you have any suggestions on how to best present/set out guidance?
6. Do the examples of good practice improve guidance?
7. Is the guidance clear as to where policy makers should go for additional help and advice?
8. Is the Small Firms' Impact test process simple to follow?
9. Do you think the Small Firms' Impact test will lead policy makers to take better account of small business issues when giving policy advice to ministers?
10. If you represent a small business, would you be prepared to take part in a focus group as part of the Small Firms' Impact test?
11. Is guidance on the Competition Test clear?
12. Is sufficient information and help provided given to estimate costs and benefits to enable policy makers to undertake a sound assessment of the likely impacts?
13. Is it helpful to combine guidance on domestic and European regulatory proposals into one document?
14. Are there other alternatives to regulation which you think should be included in the guidance?
15. Is guidance on when to carry out or not a RIA sufficiently detailed?

Source: Government of the United Kingdom. Answers to OECD survey on *ex post* evaluation of regulatory tools and institutions (2003)

95. Another example of a performance test is Hahn (2000), who - in addition to the compliance tests discussed in the previous section - subjected the content of US RIAs to two other tests: whether it was "transparent" – so that the reader could easily find what was being assumed in the analysis and could follow all the calculations – and whether it was internally consistent, so that the same assumptions were used throughout. The authors' operational tests for these criteria were the presence of an executive summary and the treatment of the discount rate. While these appear to be very basic tests, they found that only half of RIA met the former, while only 86 percent used the OMB-specified discount rate throughout.

96. A number of case-studies of outcomes of individual regulations have pointed to the critical importance of *timing* to the usefulness of RIA. A more qualitative approach could therefore focus on the timing of the incurred changes, working on the general assumption that the effects of changes in RIA as

early as possible in the regulatory process have a much more significant effect than "add-ons" at the very final stages of a RIA. Many RIAs are not initiated until after the regulatory process is well underway, and often after the preferred alternative has been selected. In this situation RIA obviously has difficulty being influential. Worse, it puts pressure on the analyst not to deliver bad news about benefits and costs, especially about the preferred alternative, leading to cynicism about the role of RIAs in the regulatory process.

97. A review of RIAs in the Netherlands in the 2001-2002 illustrates how the focus on timing can be the pivotal point for a RIA evaluation's findings and outcomes. The review was primarily based on face-to-face, open-questioned interviews. It showed that RIAs were usually carried out too late in the process to allow them to have a substantial impact on outcomes. By the time RIAs were finalised, most policy-choices had already been made. As a consequence, RIA rarely led to changes in the proposed regulation, much less to more substantial shifts, such as decisions to abandon the proposed regulation or to pursue the regulatory objective via another policy instrument. The evaluation led to an overhaul of the RIA process, breaking it into two phases. The first phase consists of a *quick scan* in which the responsible ministry establishes the necessity to regulate; identifies possible alternatives; and considers in general terms the expected effects on businesses and the environment, and the regulation's enforceability. The quick scan must also include the responsible ministry's recommendations as to which additional and more thorough tests are needed. The quick scan is scrutinised by a *Proposed Legislation Desk* (PLD) composed of staff from the ministries of Economic Affairs, Environment and Justice, leading to an agreement on which impact assessments should be carried out in full. This early involvement of a regulatory reform oversight body allows the insights from the initial RIA to be weighed by a body external to the regulator prior to their being a substantial commitment to the regulatory proposal. In the second phase, the responsible ministry carries out the required impact assessments. Before tabling a final draft regulation to the Cabinet, results of the impacts assessments carried out are presented to and discussed with the PLD. In the rare case of disagreements between the PLD and the proposing ministry, this is reported in the final RIA presented to the Cabinet.

98. Despite these negative factors, twelve case-studies collected by Morgenstern and Landy (1997) suggest that, even in cases where the RIA got off to a late start, it did frequently have positive effects in terms of decreased costs, increased benefits and promotion of regulatory alternatives. However, these effects were considered to be lesser than would have been the case given an earlier start to RIA.

99. As mentioned above, another type of performance test is based on a reconciliation of actual effects with those that were predicted by the RIA prior to the regulation being introduced. Studies of regulatory effectiveness are relatively common. For example, in the United States, the EPA is required to report five-yearly on the benefits and costs of its regulations, while other agencies, such as the US Geological Survey, provide more general reports of progress in improving underlying performance objectives, such as water quality. Harrison (2003) also cites Dutch and Swedish examples of ex post evaluation of regulatory impacts. However, he concludes that while studies of the effectiveness of regulation are relatively common, studies of their costs are much rarer, and suggests that government authorities have lesser incentives to study costs.

100. Moreover, it is still more uncommon to find *ex post* studies examining the actual costs incurred by regulations and comparing them to the estimated costs. Only very little actual work has been carried out on the predictability of RIAs or other *ex ante* estimates with the actual *ex post* costs. Such studies, however, can be seen as a crucial purpose of ex *post* analysis of RIA, namely to determine whether there is any detectable systematic bias in the observed differences between ex post outcomes and ex ante predictions. Clearly, from a policy perspective, the identification of such errors is a first step toward a review of their underlying cause and, ultimately, consideration of means by which the biases may be reduced or eliminated.

101. One example of such ex post studies of ex ante estimates of regulatory costs is a recent review of the predictability of administrative cost assessments made by Denmark's Business Test Panels. The Danish Business Test Panels are used to obtain information directly from businesses on the administrative burdens expected to be associated with proposed new regulations. The Business Test Panels have operated since 1996 and the system was expanded in 1997 so that three panels of 500 firms each were available. This move was expected to enhance the statistical reliability of the estimates obtained[9]. The use of BTPs had prior to the review, been at the discretion of regulators. However, a new policy proposal was to make their use compulsory for all new business regulation which has been judged *a priori* to have impacts on business above a certain threshold size[10]. The review intended to ensure the credibility and usefulness of the test panels in the context of this intended expansion of their operations. It was carried out by a consultancy engaged by the Ministry of Economic and Business Affairs' Commerce and Companies Agency. It covered three years of assessments carried out by the test panels. Quantitative statistical analysis was used to measure the actual impact of past regulations. Surveys and interviews were used to better understand test panel enterprises' approach to assessing impacts *ex ante*. The review found that the Danish BTPs estimated the realised administrative burdens of new business regulations with a margin of error ranging between 40 and 60% on average. This was viewed as acceptable in light of the inevitable complexity and uncertainties involved in *ex ante* evaluation and the purpose of the assessments. However, despite the generally positive assessment, the review led to a number of improvements in the methods employed for extrapolation of data. (Other aspects of the results of this review are discussed in the section on Function Tests below).

102. Another example of a study comparing ex ante estimates with ex post outcomes is an independent study carried out be Morgenstern et al (2000) covering environmental regulatory programmes in the United States. The study indicated that *ex ante* estimates tended to exceed actual costs, mainly due to difficulties in defining baselines and estimating (incomplete) compliance costs. The study also showed that unit cost estimates were often accurate, although for rules that use economic incentives, unit costs were consistently overestimated. Findings of the study also indicated that benefits may be overestimated. In cases of unit-cost overestimation, Morgenstern et al. argues that unanticipated technological innovation appeared to an important factor, especially for economic incentive rules.

103. Morgenstern et al's findings support the view of some proponents of regulation that RIA tends systematically to over-estimate regulatory costs. A key reason advanced for this view is that it tends to adopt a static approach and ignore the often crucial effect of technological innovation in reducing actual costs. Another argument often put forward to support the view that RIAs often overestimate costs is that economies of scale will be reaped in the production of whatever product or service is required due to a regulatory standard, and that these will result in, often substantial, reductions in both real costs and market prices.

104. However, a counter-argument against these assertions of a tendency to over-estimate costs is that these technologically based reductions in costs are substantially the result of a diversion of research and development expenditures into a particular area that has been rendered potentially economically attractive by the implementation of the regulatory requirement. Such diversion of R & D resources necessarily has an opportunity cost attached to it – i.e. that of the foregone productivity of such R & D resources if committed to an alternative use. The size of these opportunity costs is clearly not susceptible to measurement, but they are clearly likely to be important in terms of a consideration of the overall impact of regulation. This example perhaps serves to point to the limits to RIA. However, in doing so it also suggests that any perceived tendency toward over-estimation of costs may, itself, be over-estimated.

---

9 See *Regulatory Reform in Denmark.* OECD, Paris, 2000, p155.

10 The threshold is that the impact would be at least equivalent to around 2000 hours of paperwork

105. Morgenstern et al's finding that benefits also tend to be over-estimated - as a result of a tendency to assume full compliance with a regulatory standard – may imply that systemic bias in net benefit/cost estimation may not, in fact, be significant. Thus, the predictive ability of RIA may be somewhat better than a first glance at the Harrington results suggests, at least in the specific sense that the predicted balance between benefits and costs may be broadly accurate. However, it does of course not constitute an argument against the need to consider RIA methodological approaches in order to try to minimise the observed inadequacies in analysis. As the OECD has previously argued (see, eg. OECD 1997, p216), the dynamic impacts of regulation are notoriously difficult to predict accurately and incorporate in RIA, and this is an area in which ex post analysis may have much to contribute. It is at least theoretically possible to consider ex post analyses as providing "feedback loops" in a process of refinement of RIA methodological guidance that aims particularly to improve the ability of the tool to deal with the dynamic aspects of regulatory impact.

106. It should be noted that the question of how well RIA takes dynamic impacts into account can be considered from more than one perspective. The above discussion presents the view that costs can often be over-estimated due to failure to account for technological shifts. However, a broader concern in terms of the consideration of regulatory impacts in the dynamic context is that the market distortions which regulation frequently introduces – and which is often indirect in nature and ill-accounted for in RIA – are likely to become progressively larger over time and capable of overwhelming the benefits which the regulations initially sought to achieve. Shifts in social and economic circumstances mean that regulation departs progressively from a 'best practice" state, while review and reform may be long-delayed. The role of ex post analysis in documenting these concerns is potentially a very substantial one, contributing both to our understanding of the effectiveness of RIA and means of improving it and to our understanding of regulatory dynamics more broadly.

107. The rarity of performance tests that evaluate the predictive capacity of RIAs may be due in significant part to the lack of incentives to do so for government authorities. That is, there is little benefit for them in highlighting their RIA performance, whether positive or negative. A negative result will tend to undermine their credibility when proposing new regulation – supported by RIA analysis – while a positive result does no more than provide a confirmation of the benefits of a policy which has already been implemented and in which the agency already invests its believe. A second reason for the lack of such evaluation is likely to be the technical challenges in producing relevant data. That is, there may be substantial resource requirements involved in collecting and even analysing data on the true costs and benefits of regulatory measures while, from the agency's perspective, the resources devoted to such tasks are likely to be seen as having been diverted from core tasks. Given the lack of clear benefits to the agency or its Minister from such expenditure, this resource constraint is likely to impede substantially the prospect of such evaluations being completed.

108. The cost of such evaluations can potentially be reduced substantially by integrating data collection requirements into the design of regulation and, more broadly, giving significant ex ante consideration to the issue of what data would be required in order to conduct ex post review. However, it is clear that such reviews remain resource intensive and must be targeted effectively if they are to be both feasible in resource terms and likely to pass a benefit/cost assessment themselves.

4.3.1. Models and Methodologies

109. With few exceptions, RIA evaluation has taken little notice of the utter dependence of RIAs on models. Although sophisticated models are probably only used in relatively few RIA, these are likely to RIA where very substantial benefits and costs are expected to flow from the adoption of the proposed

regulation. The quality of models is often taken for granted and left to the discretion of the preparing agency. However, inadequate models have misled regulators into writing bad regulations.[11]

110. Harrington (2004) points to substantial potential problems arising from the use of models. Many of these arise from the fact that regulators will often need to use "proprietary" models. This inevitably means that the transparency of the RIA analysis is substantially reduced, since much of the model and/or the underlying data used is generally unavailable for scrutiny. Reduced transparency clearly has negative implications for the credibility of the RIA process and, in many cases, the acceptability of the regulatory outcome. However, the prevention of peer review of the underlying model and its assumptions, which is a likely result of the use of proprietary models, is also likely to have direct negative impacts on quality.

111. The issue of ensuring the quality of models employed is clearly crucial, and an area for RIA evaluation that probably fits best under the rubric of performance testing. The question is how RIA procedures can incorporate provisions for the vetting of models used to make public decisions, including their assumptions and underlying data sets.

112. Quality control in respect of models employed can be advanced in several ways. One is to ensure that all input data (such as specific regulatory measures, quantities, regulatory standards) are transparent and freely available, and by allowing for stakeholders to engage in the consultation process on the basis of the use of other, plausible models. If different stakeholders, using identical input data but different models, arrive at different assessments of expected regulatory impacts, it is likely that this may expose and generate a debate about potential biases in the various models. A dialogue which can embrace the issue of the sensitivity of the predicted regulatory outcomes to the models employed to estimate it is clearly one which will test quite rigorously the robustness of the regulatory process and the likelihood that the regulation will be welfare-enhancing.

4.4. Function tests

113. The idea of evaluating RIAs presupposes that RIAs make a difference, that is, that the outcome of regulatory processes is in some way different from what it would have been in the absence of the RIA. Clearly, this is the working assumption underlying all of the regulatory quality tools and institutions considered in this report. However, while the compliance and performance tests cited above can tell us much about the quality of RIA, they say nothing about the extent, if any, of its impact on actual regulatory quality outcomes.

114. Function tests focus on how RIAs have changed the outcomes of the rule-making process and improved regulatory quality. Four different forms of RIA function tests can be identified[12]. These are:

- tests focussing on the frequency with which initial regulatory proposals are revised;
- tests using audit trails in relation to newly raised options;

---

11 An example of an egregious case in the U.S. noted by Harrington (2004) was the "Enhanced Inspection and Maintenance" (I/M) rule for motor vehicles in 1992. This model adopted very optimistic assumptions regarding identification of high-emitting vehicles, ability to find cheaters, and expense and effectiveness of vehicle repair. It led to cost-effectiveness estimates of well under $500 per ton of VOC and NOx emissions. Other independent estimates were near $5000 per ton, which turned out to be pretty close to the actual results observed in implementation.

12 While largely distinct, there is necessarily some degree of overlap at the margins between some of these different types of function tests.

- tests highlighting the difference between initial regulatory proposals and final regulations in benefit/cost terms; and

- tests of the effect of RIA on the administrative/regulatory culture.

115. Probably the most basic form of function test for RIA is the collection of data on the *frequency* with which initial regulatory proposals are either modified or abandoned as they progress through the process. Applying such a test effectively assumes that the modifications made (or the abandonment of the proposal) are wholly or largely attributable to the impact of the RIA process in bringing rigour and transparency to the analysis. In fact, other influences on the policy process are likely to have some influence, so that RIA will constitute only a part of the reason for any observed changes. Nonetheless, the RIA discipline is typically brought to bear on most or all substantial regulatory proposals, while other factors may be expected to intervene less frequently. This, plus the nature and extent of the specific disciplines imposed by RIA suggest that it would be expected to constitute a large part of the reason for observed changes in regulatory proposals. Moreover, a part of the impact of RIA is likely to be unobservable in practice: regulators may in some cases not put forward regulatory proposals because they are conscious that they are unlikely to survive the scrutiny that a RIA process will bring to bear (anticipated reaction). The impact of RIA in preventing poor regulations at the earliest stages of development may well be highly significant, but is difficult, if not impossible, to observe directly.

116. A number of attempts at measuring this aspect of RIA outcomes have been made. Results, in terms of the frequency of changes, show wide variability, possibly in part a result of definitional and methodological differences, but likely also to reflect differences in the quality of RIA requirements and their application in practice. The general conclusion, however, is that outcome measures of this sort can be developed in practice. Examples include:

- In 1999, the OECD[13] noted that 60 per cent of draft regulations were amended during the process of RIA-based review by the Office of Management and Budget.

- Formsma (1998, p220) reports that in the 1995/96 period in the Netherlands, some 17 per cent of proposals subjected to RIA were either modified or abandoned.

- The OECD (1999b)[14] calculated that 9 per cent of regulations in the Australian State of New South Wales were modified or abandoned during the RIA process, and reported that a similar review in the state of Victoria had earlier shown a rate of around 20 per cent.

117. Tests of the rate of modification of regulatory proposals may constitute good basic indicators that RIA is affecting outcomes, but cannot provide data on the importance of these changes in practice. A more ambitious approach is to *measure the differences between initial regulatory proposals and final regulations in benefit/cost terms*. This, at least theoretically, allows a monetized estimate of the impact of RIA to be generated.

118. As suggested above, a fundamental difficulty with this kind of functional test is that the baseline, or "counterfactual" is difficult to conceptualize. Even in the absence of formal RIA requirements, it is highly likely that some analysis of the effects of regulation would be undertaken. Put alternatively, a simple measurement of the difference between the regulatory proposal initially advanced and the final

[13] Regulatory Reform in the United States, p153.

[14] Report by the Public Management Services of the OeCD on Regulatory Impact Assessment in New South Wales. Regulatory Reform Committee, Parliament of New South Wales, Report No. 18/51, January 1999, p36.

outcome does not constitute an accurate measure of the impact of RIA, since the initial proposal is, in any case, likely to be subjected to modification in the course of its development. Moreover, even in the presence of RIA, other aspects of the policy development process may also be said to have played an important role in shaping the final regulation, leading to questions as to the "attribution" of the changes to RIA or to other elements. In this respect, an analysis comparing initial proposals with final regulations risks over-stating the impact of RIA.

119. The "audit trail" approach can be seen as a variant on the approach of measuring the frequency of changes in initial proposals. It is one which avoids the above-mentioned vulnerability of the simple measure of the frequency of changes, in that it focuses specifically on the treatment of suggested changes to the regulatory proposal that have been made by stakeholders during the RIA process. Thus, audit trails review the handling of such suggestions by the regulators (and, implicitly, by regulatory reform authorities responsible for quality assurance in respect of the RIA system). Yarrow (2004) argues that there is likely to be a systematic bias against newly raised options brought forward during the RIA consultation process:

> In some ways, regulatory policy development is like a R&D process. There may be a number of ideas and options at the outset, but cost considerations lead to the sequential 'closing out' of what look like less attractive options, in order to devote more resources to the development of other, favoured options. Ideas and evidence introduced other than the very early stages may have difficulty getting a hearing, because, for example, they point in directions that might already have been closed down.

120. It is because of the likely existence of such a bias that the adoption of audit trails can be expected to be productive: they increase the pressure on regulators to deal openly with such proposals and require them to be able to justify their treatment of them. Yarrow notes that such approaches are likely to be resource intensive, but argues that, if sufficiently narrowly specified (e.g. being focused solely on the treatment of discrete new options raised during consultation on RIA documents) they can feasibly be used as a supplement to other evaluation methods.

121. Perhaps as a result of the difficulties of attributing changes to initial proposals to the RIA process, there has been remarkably little analysis of the effect of RIA characteristics on regulatory outcomes to date, despite the extensive use of RIA in many OECD countries over an extended period.[15] Where such studies have been undertaken, they have differed substantially in their approach and scope. An early example was a study of 15 regulations for which RIA were prepared, undertaken by the United

[15] One relevant though somewhat dated study of the impact of regulatory documents noted by Harrington (2004) is Magat et al. (1986), which examined the effect of the quality of regulatory support documents generally on the outcomes of the Effluent Guidelines regulatory process during the 1970s. Two documents were examined: the "development document" and the "economic analysis." The former gave the technical information on the industry, its technological options for wastewater treatment and the one identified as the basis of the regulation; while the latter assessed the effect of the proposed regulation on costs, prices, profits, plant closures and unemployment. The authors used a fairly elementary definition of document quality; namely, were the numbers consistent? Did the report leave a trail that a careful reader could follow to connect the input data with the outputs, i.e. the estimated effects? What they found was that document quality, defined in this simple way, made a substantial difference in how much the Agency changed the regulation during the rulemaking process. The more incoherent the document, the more the effluent standards changed. For example, when the development document failed their quality test, the promulgated BPT standards were made 33% less stringent for biochemical oxygen demand (BOD) and 44% less stringent for total suspended solids (TSS) than the proposed standards. The findings imply that document quality can affect the regulatory outcome. (It is possible, however, that poor documentation simply indicated an industry that was both difficult to regulate and difficult to characterize in a technical report.)

States EPA in 1987. This study reported highly quantified results: The EPA concluded that the costs of its proposed rules had, overall, been reduced by $1 billion as a result of the RIA conducted.[16] The cost of the conduct of RIA on these 15 regulations was reported as $10 million, implying a benefit/cost ratio of the RIA itself of 1000:1. Thus, it was concluded that RIA had been highly effective in relation to this sample.

122. The fourth type of function test focuses on the effects of the RIA system on the administrative/regulatory culture, that is how and whether RIA are instrumental in instilling a greater appreciation and understanding of the benefits of the RIA process, and thereby encouraging a proactive rather than reactive use of the RIA as a policy development tool. Underlying the use of RIA is the presumption that it is only when the fundamental logic of the process is fully accepted and adopted in practice by regulators that its full benefits will be attained. It is generally assumed that this is a process of cultural change that will necessarily be achieved only in the medium to long term. Thus, this type of function test can be seen as measuring directly the extent to which this long-term is occurring in practice. Such a test is, perhaps, the most important form of functional test.

123. A review of the Canadian RIA system, published in 2000, had as its objective to assess the effect of RIA in instilling discipline in analysis and affecting decision making by providing certain types of information (Delphi Group, 2000). Based on an in-depth review of six regulations and interviews with stakeholders the study concluded that RIA requirements had changed the decision-making process in Canada: "More attention is paid to alternatives and costs and benefits than appeared to exist when the requirements were instituted fifteen years ago. Officials were sensitive to RIA requirements and departments had systems in place to consider regulatory options and costs and benefits. Resources were being devoted to these activities and a core of expertise was available in several departments."

124. Although tests of RIAs impact on the administrative/regulatory culture are rarely the sole objective of RIA evaluations, broad and open survey-based evaluations have often been able detect and investigate such changes. In a review of its Regulatory Impact Statement system, New Zealand noted that the introduction of RIA framework had begun to change the structures and practices involved in the development of regulatory proposals, for instance by encouraging departments to use the RIA framework as a policy development tool rather than seeing it as a retrospective transparency tool. Also, in the above-mentioned evaluation of the performance of the Danish Test Panels, it was found that law-makers attached little significance to specific estimates of regulatory impacts in the RIA statements. (Rather, they were looking for recommendations that could be make regulations more attuned to the administrative set-up of businesses). One of the consequences of this finding was to improve the presentation of RIAs to law-makers, with a greater emphasis on the presentation of total cost estimates.

125. This type of function test is necessarily largely qualitative in nature and is in substantial measure subjectively applied. However, while these factors necessarily pose difficulties in terms of interpretation of results and in terms of the comparability of results over time and between different areas, the test is potentially more fruitful than other function tests in that it is a more direct measure of the extent to which the fundamental goal of RIA is being achieved.

4.5. Conclusion

126. RIA has become much more than a technical add-on to the rule-making process. As Harrington (2004) argues:

> "At one time it was customary to regard the RIA as a document that accompanied the release of a proposed regulation. It is apparent from a recent survey of RIA practices among OECD

[16] See Regulatory Reform in the United States. OECD, Paris, 1999, p154.

> countries, however, that the RIA has come to mean more than that. (OECD 2003) It is nothing less than an amended rulemaking process, one that contains not only all the characteristics of rational public action—careful statement of objectives, formulation of alternative approaches, consideration of benefits and costs of those alternatives, and allowing those considerations to affect the outcome—but also nontrivial outreach to and consultation with the interested public. Thus "RIA evaluation" has almost become synonymous with the evaluation of regulatory procedures and outcomes."

127. Given the central role that RIA itself has assumed within the rule-making processes of most OECD Member countries, and the lengthy experience which many countries now have with the use of this tool, the relatively low level of evaluation activity apparently being undertaken is likely to constitute an important barrier to the dynamic evolution of the RIA tool. Examples cited above show that significant improvements to RIA models have resulted from the conduct of such evaluations, yet they remain ad hoc and infrequent. Given this background, and the fact that regulation itself is increasingly subject to regular review requirements, there may be merit in adopting similar, regular review requirements in respect of the RIA tool itself, as a means of ensuring that options for benchmarking and improvement are provided.

128. The above discussion has identified examples of each of the three basic kinds of evaluation – compliance tests, performance tests and function tests – in the RIA context. While data are scarce, it does not appear that there is a strong tendency toward the adoption of any one test over the others, although compliance tests seems to the most frequently applied test.

129. Section 3 suggested that the focus, in applying different kinds of evaluation tests, should arguably shift with the stages in the implementation of the regulatory tools. For example, in the early stages of RIA implementation, a key focus must be on achieving a high level of formal compliance, while in a more mature RIA system the focus should shift toward the question of what effects on policy outcomes were observable. However, the above discussion provides little evidence that this approach has been adopted in practice. For example, the United Kingdom and Mexico, which both have relatively recently introduced RIA processes, have undertaken significant performance evaluation, while the United States, which has a quarter-century long history of RIA, appears to demonstrate a continuing focus on large-scale compliance testing (albeit frequently in an extra-governmental context). That these tests continue to demonstrate significant problems with specific compliance could give weight to the speculation that a shift toward more function testing would potentially yield the strategic insights into the failings of the system that would be needed to assist in remedying them.

130. The examples in this chapter may suggest that the approaches to RIA evaluation taken in practice are more determined by opportunity and the particular focus of the evaluator than a longer-term strategic view. One point which seems clear, is that there is little or no sense of integration of the use of the different kinds of RIA evaluations. Such an integration could potentially yield important benefits by allowing links to be drawn between different aspects – for example the observed degree of formal compliance and the extent of the impact on policy outcomes – particularly if evaluations can be conducted over a period of time.

131. The question of *who* conducts the evaluations also yields a range of answers. In the United States, private groups are substantially involved (e.g. the work of Hahn). In many cases, the work is undertaken by regulatory reform institutions, as might be expected. In others, there is an apparent link to the wider audit and review functions of government (e.g. the involvement of the UK National Audit Office). Advantages would seem to attach to the involvement of each kind of body. In the case of regulatory reform institutions, it can be expected that evaluations would be informed by a clear understanding of the nature of the regulatory quality tool, the practical constraints in using it in practice and the importance of the tool within the wider regulatory quality context. All of these might be expected

to assist in yielding a high quality evaluation with the potential to improve the design and use of the RIA tool. In the case of external bodies, the key advantage would seem to be the transparency of the review activity and the degree of pressure placed on governments to respond adequately to criticisms and identified shortcomings. Where broader audit and review bodies are involved, the evaluation of the RIA tool is placed in the broader evaluation context, while it is possible that their key role at the centre of government may also have a positive impact in ensuring that "feedback loops" are effective, so that the RIA tool's design is improved in response to identified problems.

5. CONSULTATION MECHANISMS

5.1. Introduction

132. Any evaluation of a regulatory quality tool must have, as its starting point, an identification of the objectives that the tool is intended to serve. In the case of regulatory consultation and communication mechanisms, this identification of objectives is a complex task in itself. As previous OECD work has documented (OECD 2002) consultation mechanisms can have several, inter-related objectives. At a fundamental level, consultation can be seen as having two "high level objectives":

- Supporting democratic values; and
- Improving the overall effectiveness and efficiency of policy.

133. However, these two basic objectives can be broken down into a number of subsidiary elements which must be clearly distinguished as the basis for evaluating whether consultation mechanisms are effective and efficient in practice. Serving democratic values incorporates the broad notions of transparency and accountability: that governments must be open with their constituents regarding their actions and accountable for the purposes and results of those actions. Provision of information by regulators as part of the consultation process is part of this dynamic. Serving democratic values also encompasses the concept of legitimacy: regulatory requirements are likely to be seen as more legitimate if affected parties have had the opportunity to play a role in their development, including provision of data and opinions on the likely effects of the proposal. Objectively, transparent and consultative processes reduce the opportunities for regulatory failures caused by regulatory capture and, more generally, provide stronger incentives to regulators to develop effective regulation.

134. This concept of legitimacy, in turn, has implications for the second high level objective – that of effectiveness and efficiency. It does so to the extent that a greater degree of perceived legitimacy will tend to enhance the degree of voluntary compliance with a regulation. This reduces the need for enforcement activity to be undertaken and is likely to increase the overall level of compliance (given the inherent limits to formal enforcement as a mechanism for securing compliance).

135. Consultation seeks to serve regulatory efficiency and effectiveness in several other ways. Increasingly, as noted in the previous section, consultation is integrated with RIA processes. Here, the objective is to enhance the informational basis for decision-making by seeking relevant data directly from stakeholders. This contributes to efficiency, since it is likely to constitute the lowest cost means of data collection and hence expand the total data-base feasibly able to be amassed. Second, open consultation processes are likely to improve the quality of debate, by drawing more participants into the process and providing for more intensive and fruitful interactions between them. This process should contribute to a higher quality of analysis of proposed options and the data provided and thus contribute to better decision outcomes.

136. In addition to objective data, of the sort considered above, consultation can have the objective of providing information on the subjective *acceptability* of different regulatory options. This concept of acceptability reflects particular values of the stakeholder groups and will in turn provide crucial information to regulators on the likely degree of compliance with regulation (see above) and on the extent

to which the public – or sections of it who are intended to benefit from regulation – will see particular proposals as meeting their underlying objectives. For example, market based solutions to a particular problem, such as pollution, may be objectively effective, but have limited acceptability to local residents in particular circumstances.

137. By informing stakeholders, consultation and communications strategies also serve the objective of reducing regulatory uncertainty. That is, regulation becomes more predictable and the costs of complying with it are reduced if it is the outcome of an open process with a logical progression from regulatory objective, through data collection to regulatory outcomes.

138. Reflecting this diversity of regulatory objectives, the specific consultation/communication tools employed by government also vary. As previous OECD work indicates, consultation and communication tools include public meetings, more restricted face-to-face consultation, notice-and comment procedures, circulation for comment various kinds of reference group arrangements, and publications of registers of new or proposed regulations. There can also be as a range of variations on these instruments reflecting factors such as the stage of the consultation process at which the instrument is being used. For example, the extent and type of the information provided during a notice-and-comment procedure may vary substantially according to whether it is being undertaken to inform initial decision-making on whether to undertake regulatory action, or to identify and assess specific regulatory choices. The requirements are different again when the focus is on communication of regulatory decisions and compliance requirements once decision-making has been completed.

139. Evaluation of consultation tools is inevitably made more complex and difficult by this diversity in their nature and purposes. Nonetheless, an analysis of evaluation tools for consultation and communications mechanisms can be conducted within the broad framework identified in Section 3, above, as discussed in the following sections.

5.2. Compliance tests

140. Consultation programmes are frequently subject to relatively detailed procedural and content requirements, established via legislation, subordinate instruments or government policy statements of various kinds. This may be so particularly where there is a high level of integration of consultation and RIA requirements and activities. In other, more consensus driven, political cultures, the specific requirements may be left largely to the discretion of the regulatory agency, but the exercise of that discretion will be substantially curtailed in practice both by the general objectives and requirements specified and by customary expectations regarding the requirements applicable to such processes.

141. Conducting compliance-based evaluation is clearly more feasible where there are detailed standards in place. While the specific matters dealt with in government guidelines, policies or laws governing consultation necessarily differ, matters that are commonly included and are, at least to some extent objectively verifiable in a compliance testing context include:

- Stage of the regulatory process at which consultation is to be conducted;
- Who is required to be consulted (e.g. the broad public, particular specified stakeholder group(s));
- What information is to be provided by the regulator;
- What is the minimum period of consultation required;

- What responses are required of the regulator (e.g. is a formal response to specific consultation inputs required);

- What is required to demonstrate responsiveness to consultation inputs generally?

- Are publications requirements met (e.g. communication of regulatory decisions and compliance requirements).

142. Countries have reported only few examples of compliance tests of their consultations procedures, despite the relatively uncomplicated approach and low-resource requirements to carry out these evaluative tests. However as illustrated by the Norwegian and Japanese examples below, findings of such reviews may reveal ample scope for improvements. Two reviews of the Norwegian government's consultation practices showed a low level of compliance with consultation time limits set out in government guidelines. In 1995, nearly 80% of the consultations were completed within a shorter time frame than the three months principal rule, and more that 25% were completed under the minimum time frame of six weeks. In 1997, only marginal changes had occurred: More than 75% of consultations were carried through within a timeframe of less than three months; a little more than a fourth of the consultations were completed in less than six weeks. In follow-up to the surveys, the Ministry of Labour and Government Administration distributed information about the consultation requirements to all ministries, and organised information meeting in order to increase awareness on these obligations (OECD, 2003). In a similar review of consultation mechanisms in Japan, a survey from August 2003 showed that only about half of public comment periods in 2002 were more than the recommended 30 days (OECD, 2004).

143. As noted above in relation to RIA, compliance testing of consultation essentially tells us whether regulators are demonstrating a high level of compliance with the formal requirements of the process, but can tell us little about the quality of their compliance activities, much less the actual outcome in terms of impacts on regulatory decisions. However, compliance testing is of substantial importance in relation to objectives relating to supporting democratic values by providing a more open regulatory process. As well, the relatively "objective" nature of the judgements to be made in compliance testing means that there is a greater possibility of implementing testing on a fairly wide scale.

5.3. Performance tests

144. Performance testing in relation to consultation embraces a range of quality indicators relating to the documents provided as the basis for consultation input and the actions of the regulator in managing and responding to inputs. A range is discussed below.

a. Formatting

145. One fundamental aspect of performance is the requirement that the party being consulted understands the nature of the regulatory proposal or question and the major decision-factors that will come into play. This requires, firstly, that the appropriate information is published in consultation documentation and, secondly, that it is presented in a form that is easily intelligible and facilitates the provision of relevant responses. Excessively complex or opaque consultative documents can yield a number of substantial negative impacts. In particular, they can lead to the emergence of "insider/outsider" dichotomies, where the voices of incumbent groups, who have made substantial investments in consultation processes, effectively gain excessive weight at the expense of those attempting to participate for the first time, or without adequate support or expertise. Such a dynamic is likely to both encourage, and result from the mechanics of regulatory capture – that is, the tendency of regulators in particular circumstances to see their incentives as being aligned with particular incumbent producer groups. A result

is that sub-optimally large amounts of resources can be devoted to the management of regulatory and government affairs.

146. Performance testing in this respect can be conducted via controlled experiments, where participants who have no experience of regulatory consultation in the context under review would be taken through exercises designed to discover how easy it was to discover and comprehend relevant information relating to the consultation process – such as would allow them to respond effectively to the documentation. Alternatively, a review of the format and content of consultation documents could be conducted in terms of a pre-determined set of criteria based on an *ex ante* view of what information is required, in general terms, for effective participation. An indicative list of such criteria could include the following:

- To what extent does a given document clearly signal the context of relevant past and ongoing consultation processes, and of regulatory and market developments?

- Are those powers and responsibilities of the consulting body which are relevant to the issue being considered clearly set out in the document and/or referenced in an accessible way?

- Are the objectives and the scope of the consultation process clearly defined?

- Are the possible outcomes of the consultation process highlighted (e.g. are potential regulatory outcomes identified)?

- Are the steps that would be necessary in order implement different types of outcome highlighted, with an indication of potential timescales (e.g. is new legislation necessary, and if so what implications would that have for relevant timescales)?.

- Is a timetable provided that sets out when it is planned that the different stages of the consultation process will take place?

- Are commitments made with respect to what respondents can expect it they make a submission -- for example, setting out how will responses be reviewed (see further below on processing of responses). Will there be a summary of issues raised by respondents in a subsequent published document? Will responses be provided to material issues raised by respondents?

- Does the document identify parties who are expected to be materially affected by the matters being considered (including by any potential subsequent regulatory changes)?

- Are contact details provided in order that interested parties can seek clarification and/or further information concerning the document?

- Is a summary of the document provided, which adequately reflects the scope and the likely relevance of the document to a reader?

b. Targeting

147. While the issue of formatting, discussed above, relates to the effectiveness of consultation processes, the question of whether consultation effort is appropriately targeted relates to its efficiency. That is, properly targeted consultation efforts reduce the resources required to be devoted to this activity both by regulators and, in particular, by stakeholders.

148. One aspect of this concept of targeting is that of regulators exercising strong disciplines over the nature and quantity of the information that they publish. The prospect of "information overload" is real and substantial: for stakeholders, receiving large quantities of often irrelevant information greatly increases "search" costs, as they must sift through this material to identify that which is relevant to them and to which they will choose to respond. For regulators, the problem is smaller but nonetheless real, since publishing large quantities of information is itself resource-using and only justified by a real prospect of better consultative responses. Another aspect of this issue of targeting relates to the type of consultative tool used: while notice and comment processes yield a high level of assurance that all potentially interested parties have an opportunity to be involved, it also can yield a high degree of "information redundancy" for many parties.

149. Secondly, communications that are undertaken in a consistent manner over time are likely to reduce the costs to stakeholders of "scanning" to ensure that they are aware of relevant consultative material. This is particularly important in relation to broad-scale consultative tools, like notice-for-comment procedures. Publishing consultation notices in known places, perhaps on known days, can simplify the task of monitoring regulators' outputs, increase the probability of reliable identification of relevant consultation opportunities and improve the timeliness (and hence practicability) of response.

150. Performance testing in relation to these consultative/communications criteria would focus directly on indicators such as the frequency and timing of consultative documents, publications policies (i.e. what media are used for dissemination) and the range of stakeholders alerted to particular issues (i.e. where consultation is of the targeted type, such as circulation for comment, is there a good match between those to whom materials have been circulated and those with identifiable and substantial interests in the matter at hand? A performance indicator of the success of this type of targeting might be the rate of response from stakeholders. This is a relatively low cost form of *ex post* evaluation, if conducted contemporaneously with the consultative process. On the other hand, the results of such an indicator can be ambiguous: a low response rate may indicate general satisfaction with the nature and direction of the regulatory proposals, but may equally be a reflection of cynicism as to the value of engagement in the consultation process. This argues for the use of the indicator in conjunction with others that can provide additional information that would reduce or eliminate the ambiguity. An example of a closely related indicator in this case would be a qualitative breakdown of the response rate data to determine what stakeholders were represented. Similarly, qualitative judgements about the "quality" of the input – in the sense of the extent to which efforts to provide useful and sophisticated responses had been undertaken.

c. Response profiles

151. A direct measure of the effectiveness of consultation efforts is the number and quality of responses received from stakeholder groups. A larger number of responses will, ceteris paribus, tend to indicate that stakeholders are engaged with the consultation process and is thus a likely indicators that they have confidence in the potential effectiveness of expending these resources in this pursuit. Measures of the quality of responses could include attempts to quantify the number of alternative proposals put forward, or the amount of factual information gleaned from stakeholders. Measures of response rates and quality can also be made at a disaggregated level, to determine whether the effectiveness of communication with particular stakeholder groups is adequate and so assist in determining if and how to revise future consultation strategies.

152. In Japan, the Ministry of Public Management, Home Affairs, Posts and Telecommunications (MPHPT) within the framework of the Government Policy Evaluations Act carries out annual surveys of ministries' compliance and performance on a number regulatory tools, including consultation practices. The surveys also systematically cover response profiles to all regulations subject to public comment. (In 2003 one third of all consultation documents received no comments; around 40% received between 1 and

10 comments; 25% received between 11 and 100; and around 5% received more than 100 comments.) Results of the surveys are made publicly available, and serve as an appropriate platform to monitor and communicate performance on consultation practices and other regulatory tools.

d. Adequacy of reasoning

153. A further performance indicator for consultation mechanisms would focus on the adequacy of the reasoning supplied in consultative documents, both in terms of the justification of initial proposals and in terms of subsequent publication of responses to inputs received from stakeholders. The latter is the more obviously relevant indicator, since it is a good partial measure of responsiveness to stakeholder input. Unexplained neglect of substantive stakeholder inputs is clearly an important indicator of a consultation process that is failing to function effectively as an aid to regulatory quality.

154. The opportunity to undertake performance testing of this kind is generally increasing with the trend for greater documentation of regulatory decisions and processes to be required. This is clearly an area in which the increasing links between RIA and consultation processes is also highly relevant, as an integrated RIA and consultation process often sees a RIA type document released as the key informational element of the consultation process. Thus, there is a clear potential for "crossover" between performance testing in this area and in relation to RIA itself.

155. Checks on adequacy of reasoning could be conducted by an internal "editorial board" within the regulatory agency, by an external assessor, or by a panel that mixes the two inputs. Arguably, the appropriate role for an internal editorial board within this context would be as an *ex ante* quality control mechanism before documents are publicly released.

e. Evidence checking

156. Closely related to the notion of testing for the adequacy of reasoning in respect of responses to consultation inputs received is the concept of "evidence checking" as a performance test. That is, the performance assessment would examine the weight and emphasis given to the views received and evidence provided during the consultation process in the development of the final decision and the explanatory material that accompanied it. Such review actions could include determining whether there were meetings to evaluate responses, whether decision makers review responses directly, or rely on summaries, or whether advisory bodies are involved in assessing responses. Of particular importance, the assessment could determine whether evidence that was unfavourable to the regulatory proposal was presented and adequately rebutted or placed into context within the justification given of the final decision.

157. A more detailed variant of this process would focus not simply on the links between final decision-documents, and the underlying analysis of the regulatory decision contained in them, but also seek to audit the internal "processing" of the consultation input. Such an audit could be based on a checklist of questions such as:

- Are there recorded meetings within the regulatory agency to evaluate responses?
- Do actual decision-makers review responses, or summaries of responses?
- Are external advisors involved in assessing/reviewing responses?

158. There may be substantial reasons why evidence provided does not yield substantive change to regulatory decisions, even when it is assessed objectively and fully. As noted by Yarrow: *"…even assuming effective regulation, it is to be expected that the information supplied by an interested party will*

not typically affect eventual outcomes, either because the information was already known or because the individual submission, although it provides new information, does not affect the perceived overall balance of advantages of a particular course of action".

159. Given this fact, it is particularly important that the actual treatment of inputs is made transparent precisely in those cases where there has been no clear impact on outcomes. From the point of view of the stakeholder, continued participation in the process is likely to be crucially dependent on the perception that such inputs can, at least potentially, have an impact on outcomes. Thus, the ability of the consultative process to demonstrate that it has weighed seriously the inputs received is a key performance indicator that is likely to be predictive of its longer-term quality and effectiveness.

160. Such performance tests are likely to be relatively resource-intensive in nature, requiring the assessor both to obtain a high level of familiarity with the consultation inputs and to undertake a tracing of the links between them and final decision documents. It is also a test that necessarily requires subjective judgements and is open to interpretation. Given these factors, such a performance test would be of limited application in practice. It might, for example, be used as an intensive and targeted quality check that would be applied to a small number of far-reaching policy decisions and/or those which had had controversial or extremely broadly engaged consultation processes. This kind of "case study" approach could be seen as a periodic "check up" process, which might signal the need for a wider review of consultation processes were the results to be clearly negative[17].

f. Review of internal regulatory costs

161. As a further test of regulatory efficiency (rather than effectiveness, as are most of those considered above) it is possible to consider an accounting of the internal regulatory costs of the agency responsible for a particular consultation process. Such an exercise clearly relies on the ability of the reviewer to adduce cost "benchmarks" against which the results can be measured. This process will itself pose some difficulties, since consultations will vary widely in terms of the size of the stakeholder group, the consultation tools most appropriate for use, etc. However, a disaggregated accounting of the costs involved, identifying costs at the level of individual consultation-related activities, could potentially provide the basis for a "diagnostic" approach to the agency's use of consultation.

162. Such an exercise might also be used to highlight the different cost implications of different kinds of consultation and provide input that would assist in directing future consultation efforts at a government-wide level. This test can be equally applied for other regulatory tools and institutions.

163. However, some significant difficulties are clearly apparent, including the practical ones of ensuring that consistent cost definitions, and approaches to attribution of costs, are used. The high importance of issues of comparability – and the need to compare results to achieve a benchmarking – suggests that this tool is appropriately applied from a centre of government position, rather than via regulators themselves or external consultants engaged by individual regulators.

5.4. Function tests

164. Functional tests are defined in Chapter 3 as those that indicate directly the impact of the regulatory tool or institution on regulatory outcomes and regulatory quality. Three main function tests have been identified in relation to consultation strategies: the incidence of regulatory changes in response to consultation processes and the use of surveys. These are discussed in turn below.

[17] Note that, conceptually, these tests can be considered also to have elements of an outcome based test.

a. Audit trails of new options raised in responses

165. The previous sections highlight the difficulties and subjective judgements that are likely to be required in attempting to determine how responses have been to evidence presented during consultation. Such difficulties necessarily tend to detract from the clarity of any results of such performance tests and thus cast some doubt on their practical utility. A variant of the above performance tests that is less prone to this problem is the conduct of *audit trails* in relation to new options raised in consultation responses. This test is effectively equivalent to that proposed above in respect of RIA. The focus is on the question of whether a specific regulatory alternative that has been identified is analysed adequately and compared objectively to the original proposal.

166. Conceptually, such a performance test can be considered potentially very powerful in terms of its diagnosis of the quality of the consultation process. This is because there are obvious reasons to assume that many regulators will be biased against a newly proposed alternative. These would include:

- A degree of "commitment" to the regulatory proposal, arising from their role in it's development and its promotion within government as the likely preferred option;
- An associated unwillingness to concede that others beyond the professional policy apparatus have identified a superior policy response;
- The possibility that there is a degree of prior political/administrative commitment to the draft regulation on the part of government more generally, leading to a reluctance to change;
- Unwillingness to bear the additional administrative burdens and the costs involved in delaying the regulatory process, which would be associated with moving to a new proposal.

167. Thus, audits of particular regulatory activities can be used to check for failures of the regulatory process of this kind. These would be *ex post* and infrequent in nature, since the size of the task – involving identification of relevant options proposed and a review of their treatment through the remainder of the process – is substantial. However, the relevance and value of the process would be that a failure in a specific case would potentially identify a systemic problem, while the detailed level of analysis undertaken is likely to provide a source of important information regarding the nature of the failure that occurred. This, in turn, is likely to be of operational value in considering appropriate ameliorative measures.

b Incidence of policy changes

168. This test is also effectively equivalent to that proposed above in respect of RIA. That is, the focus is on measuring directly how frequently, and to what extent, the input received during consultation changes policy outcomes. The underlying presumption is clearly that such changes will be in the direction of improvements to regulatory quality. However it is also possible that a poorly functioning regulatory process will result in consultation inputs reducing the quality of initial regulatory proposals. This is a particular danger where consultation is unbalanced, or other forms with restricted access to the process.

169. In such cases, there may be a predominance of the voices of particular vested interests, or incumbent groups, among the inputs received by regulators during consultation. These groups may then use the consultation process to argue for changes that are consistent with their own self-interest, and such arguments may be accepted by regulators if there are not sufficient alternative views being represented. This dynamic is, of course, related to broader issues of regulatory capture and/or to problems of regulators not possessing sufficient resources and expertise to allow an adequate internal analysis of views received.

170. Another issue is that the incidence of policy changes is an ambiguous indicator of the quality of the regulatory consultation process. It might be considered in general that a higher level of regulatory changes in response to consultation was indicative of a well-functioning consultation process administered by a responsive regulatory bureaucracy. On such an analysis, a decline over time in the incidence of regulatory changes would be seen as a negative. However, an alternative possibility is that such a decline reflects an improvement in functioning, such that initial proposals are better developed and less likely to be faulty, or that issues are identified and dealt with at an earlier stage, before formal consultations have commenced. Similarly, it is possible at least conceptually, for there to be "excessive" flexibility on the part of the regulator, indicating a willingness to please the consulted group even at the expense of compromising the regulatory outcome from the broader societal perspective.

171. An appropriate approach to conducting such an outcome test would be to ask regulators directly to identify instances in which their proposed regulation has been modified in response to consultation inputs. Encouraging regulators to recognise such changes in a positive light may itself have a cultural benefit, in terms of undercutting what is otherwise likely to be a tendency to see such responsiveness as an admission of prior failings in policy development.

c. Surveys

172. A third direct test of the functionality of consultation processes is to ask stakeholder groups who are the target of consultation their perceptions of the effectiveness of the process. Such tests are necessarily prey to the subjective judgements of those involved – which may not only be biased by self-interest, but also founded on incomplete information. However, a critical review of such responses can be expected to identify and correct for a large proportion of such systemic biases and allow surveys to provide substantial information on the functionality of consultation. An indicative range of questions that could be incorporated in surveys of these kinds include:

- how information is received by the interested party (ie: by regular email updates or through more occasional *ad hoc* measures);
- the quantity of information that is received: there might, for example be a view, that there is too much/too little material received from the regulatory body;
- the relevance of the information that is received, and in particular whether constituencies believe that they receive information that is of little or no relevance or value to them;
- the timeliness with which communication is received: as discussed above, the timely receipt of communication can in many cases impact upon both its relevance and value;
- search costs involved in obtaining information that is relevant and of value to interested (e.g. in identifying both the types of costs involved and some attempt to quantify those costs);
- the structure or format of the information that is received: was the information presented in such a way that its relevance was clearly identified?;
- how (by what means) and when (at what time/stage of the process) was the respondent invited to make a submission in the consultation process;
- whether the regulatory agency typically publishes (on a website, for example) or allows public access to the (non-confidential) submissions received under the consultation mechanism;

- whether there is typically a discussion of the outcome of the consultation mechanism, such as a document summarising and commenting on the received responses;

- to what extent those who have participated in the consultation process have felt that their views and submissions have been given due consideration and weight by the regulatory agency when evaluating potential options and reaching a decision; and

- the perceived fairness of the consultation scheme, so as to provide an assessment of whether or not there is a perception that the consultation mechanism is non-discriminatory or biased in favour of certain interest groups

5.5 Conclusion

173. Consultation has a substantially longer history as an element of the regulatory process than does RIA. Despite this, the application of evaluative tools to the use of consultation does not seem to be substantially further advanced than is the case with RIA. This may reflect, in part, the non-emergence, until recently, of a clear set of "good practices" against which the performance of consultation processes can be benchmarked. Previous OECD publications[18] have highlighted the fact that different consultation tools and processes have different strengths and weaknesses, while consultation can serve differing objectives. As a result, a range of consultation strategies are increasingly used in combination to achieve regulatory quality outcomes. This complexity in the nature and uses of the consultation tool implies an added degree of difficulty in identifying and carrying out appropriate evaluation.

174. This chapter has identified a wide range of potential evaluative tools for use in relation to public consultation. The rich variety of these tools may itself reflect the complexity of consultation and the multiple objectives it seeks to address. The above discussion suggests that a range of these evaluation tools should be used in conjunction with each other in order to obtain a full picture of the performance of consultation initiatives. Different tools are likely to highlight performance against different objectives, or in different contexts.

175. One clear distinction arising from the above discussion is between performance in the sense of successfully engaging stakeholders in the consultative process and performance in terms of the successful integration of insights and information gained through consultation into the regulatory process. In the former case, the challenge is to ensure that key stakeholders activily take up the opportunity to participate in consultation. This can be measured through a number of performance tests, discussed in Section 5.3., which focus on the quantity and quality of participation in the process. To the extent that participation is lacking, other tests can be used as diagnostic tools. Compliance tests that test how widely consultation opportunities were notified, what materials were made available and what periods of time were allowed for response can diagnose system design problems. Performance tests measuring the targeting of consultation efforts and the response profiles can go beyond the question of compliance with formal standards to the question of the quality of system design and implementation and provide a more sophisticated view of system design issues.

176. The second element of consultation performance is that of integrating the results of consultation into the regulatory process. This requires that regulators must be open to the inputs received from stakeholders in that they and policy-makers must be willing to depart from initial regulatory proposals to the extent that superior options are identified. This chapter notes that there are significant reasons why such responsiveness may be limited in practice. Thus, a key part of a well-functioning consultation system is likely to involve high levels of transparency and accountability; that is, mechanisms whereby there are

[18] See, for example OECD (2002), especially pp 67 – 69.

external checks on the responsiveness of regulators, creating pressure for consultative inputs to be taken properly into account. Sections 5.3 and 5.4 have identified a number of evaluation tools that focus on this aspect of consultation performance, including evidence checking and audit trails. It was noted in Section 5.5 that, while the incidence of policy changes – a function test – constitutes a more direct measure of responsiveness, it is an ambiguous indicator, in that a low level of change can be evidence of sound initial consultation and early policy design, just as easily as a lack of responsiveness. The existence of such ambiguities again highlights the need to take a multi-faceted approach to the adoption of evaluation tools in relation to consultation processes in particular.

177. It is also clear that the two elements of the performance of the consultation tool highlighted here are, in fact, inter-dependent. If stakeholders are not able to perceive a sufficient level of integration of their consultative feedback into the regulatory process – that is, a substantial incidence of regulatory change arising from the consultation efforts made – their willingness to participate will fall over time, as the expected value of that participation diminishes. This effect will operate even if the other system design elements noted above – in terms of formal compliance, targeting and the like – are of a relatively high quality. This inter-dependence, or "feedback" itself serves to demonstrate some of the evaluative difficulties that are likely to be encountered. Compliance tests might show a high level of formal compliance with consultation requirements, while performance tests show a low level of participation. This may be the result of poor performance in integrating past consultative feedback into regulation, rather than the more obvious cause of system design faults limiting effective opportunities to participate.

178. In sum, the range of evaluative tools applicable to consultation is wide, with different tools able to make distinct contributions to the overall assessment of the quality of consultation and its practical results. A sophisticated combination of these tools that is also tailored to the specific goals underlying the use of consultation in a particular country, is likely to represent the most effective approach to the evaluative process.

6. REGULATORY OVERSIGHT BODIES

6.1. Introduction

179. Regulatory oversight bodies, whether part of the executive or legislative bodies or (more rarely) having an independent status, represent an institutional response to the complexity and multiple interests involved in regulation-making. In the broadest terms, their remit is to bring a strategic and "whole of government" perspective to bear on the processes associated with employing the regulation-making tool of government. The establishment of regulatory oversight bodies represents one among a number of converging trends in OECD Countries' regulatory policies. Increasing numbers of countries are establishing these units, while the units themselves are increasingly to be found in centre of government agencies, from which position they are best placed to exercise the strategic overview function and to obtain and wield the necessary political authority to ensure their effectiveness[19].

180. Little effort appears to have been made to date to conduct ex post evaluation of the role of these oversight bodies or their effectiveness in practice. This may, in part, be seen as a result of the relatively recent implementation of this particular regulatory quality institution in many countries. However, the quantity of evaluative effort applied appears to be significantly less than that applied in relation to other regulatory quality tools discussed above (i.e. RIA and consultation), even though those tools are also of relatively recent origin in many countries. Moreover, there appears to be a relative paucity of evaluation of regulatory oversight bodies even in countries that have substantial more experience with them.

181. A second likely explanation for the low level of evaluation of the activities of these bodies is that of complexity and conceptual difficulty. As with any evaluation, *ex post* evaluation of central regulatory oversight units must have regard to the purposes and objectives underlying the establishment of the institution. The potential purposes and objectives for such oversight bodies are numerous, while any given oversight body is likely to be tasked with a combination of them. Thus, designing evaluations that have regard to these purposes and objectives will be a complex and demanding task.

182. Each country and each oversight institution will likely involve a special combination of purposes, powers and impacts on regulation. Farrow (2003) notes that the objectives of a regulatory oversight body can be viewed from within one or more of the prisms of different regulatory decision-making models: rational actor models see their role in terms of enhancing the objective economic efficiency and effectiveness of decision-making, bureaucratic process models see their role in terms of ensuring adherence to statutory or administratively based process requirements, while political economy models see their role in terms of the advancement of the political interests of the Government. Thus, while their broad task is that of enhancing regulatory quality, indicators of such quality improvements may be as disparate as greater regulatory effectiveness, improved cost effectiveness, more timely regulatory responses, improved consultative procedures, greater transparency, greater compliance with parliamentary and other scrutiny requirements, or greater consistency of regulatory outcomes with stated government policy directions.

---

[19] By 2000, 23 OECD countries had established such bodies, compared with 14 in 1996. Of this number, 20 were located in either the Prime Minister or President's office or else in the budgeting agency. See OECD (2002) pp 84 – 91 for a discussion of the functions of regulatory oversight bodies and trends in their use.

183. While the country-specific combination of objectives for regulatory oversight units inevitably differs, a general characterisation of their roles would identify three broad approaches.

> ***The Advisory Role*** involves providing advice and support to regulators to assist them in complying with government policies aimed at regulatory quality assurance. This can involve the publication and dissemination of written guidance and the provision of training on topics such as aspects of regulatory impact assessment processes and techniques. It may also involve a more specific, "hands on" approach, whereby the central unit provides advice to regulators in the context of their development of particular regulations.
>
> ***The Gatekeeper Role*** involves challenging and controlling the quality of draft regulations. This function centres on the ability of the oversight body to question the technical quality of RIA and of the underlying regulatory proposals and is likely to be based on compliance with a "checklist". However, other aspects of the gatekeeper function may involve checking and enforcing compliance with procedural requirements in relation to regulatory activities, such as aspects of consultation processes.
>
> ***The Advocacy Role*** involves the promotion of long-term regulatory policy goals, including policy change, the development of new and improved tools and institutional change. This function is a strategic one, where the other two roles are more focused on day to day regulatory management functions. The advocacy role sees the regulatory oversight body as an active player in the policy formulation process, rather than as an implementer of settled policy, as do the other two roles. This active policy formulation role, particularly to the extent that it is undertaken in the public domain, can be seen as an unusual one for an element of the administration to undertake. Sometimes, the advocacy role is undertaken by an external body appointed by the government, such as the Better Regulation Task Force in the United Kingdom or the Council for Regulatory Reform in Japan.

184. Most often, central regulatory oversight units perform elements of all three of these functions. Regardless of country specific regulatory quality objectives, the purpose of *ex post* evaluation of the effectiveness of regulatory oversight units would be to investigate the contribution of the unit's activities to improved regulatory quality. As noted above, however, this is itself a multi-elemental concept.

185. A third difficulty in evaluating the performance of regulatory oversight units is the multiplicity of factors – in addition to the activities of the oversight unit – that can contribute to the outcome being measured. This issue is particularly stark where outcome type tests are concerned, with the attempt being made to measure impacts on regulatory quality directly. Clearly there are substantial numbers of players in addition to the oversight unit who have had an effect on the final outcome, such that the simultaneous activity of the parties may make identification of individual impacts difficult. An agency may interact with the central regulatory authority in many ways so that it can be difficult to separate the product of the originating agency and the review agency. In such circumstances, only the combined final product may be observable. The "counter-factual" is also necessarily difficult to establish. Moreover, data challenges are significant. Incentives to avoid recording data involve the tendency of governments to wish to "speak with one voice," objectives may be conflicting, there can be interests in obscuring the political trade-offs that occur, and some data can be relatively non-standard such as tracking text changes in documents or the nature of meetings that occur.

186. As for RIA and consultation mechanisms, the chapter organises the possible tests into three categories – compliance tests, performance tests, and function tests. However, a further classification is also used, reflecting the three distinct roles of advisor, gate-keeper and advocate highlighted above. That

is, the tests of the effectiveness of the regulatory oversight unit that might fruitfully be applied will often differ according to which of these roles is being played.

6.2. Compliance tests

187. Chapter 3 defines compliance tests as those that focus on whether formal procedural requirements are being met. Where oversight units have a formal role in regulatory quality assurance procedures – a fundamental compliance test is that of whether those procedures are being followed systematically. An example is whether draft RIA are consistently being submitted to the oversight body for review and whether they are being submitted in a timely fashion. However, this example provides an indication of the above-mentioned difficulty in attributing outcomes to the actions of one or another body. Where compliance with RIA review requirements is lacking, the responsibility is, at least in part, attributable to the regulatory agencies. However, the oversight body clearly has a responsibility to enforce compliance, while a failure to do so might be a very effective compliance test, in that it could point to the inadequacy of sanctions or authority available to the oversight unit to ensure that compliance is achieved.

188. A second form of compliance test is that of whether the oversight body carries out its review functions in a timely and systematic fashion. Thus, data on average response times where assessments or comments are required from the oversight body might be collected and analysed. Compliance testing might also consider the extent to which the oversight body has undertaken other verification-type activities – for example, assessing and reporting on compliance with consultation requirements.

189. All of these compliance tests essentially relate to the ***gate-keeper*** role of the oversight unit. Other forms of compliance tests can be envisaged in relation to the ***advisory*** role. These would include the collection of data on the quantity of training services provided to policy staff in regulatory agencies, or the advisory materials published for use by regulators and/or other stakeholders in the regulatory process. The use of compliance tests in this context is likely to be fairly limited, however, since there will not usually be detailed formal standards and requirements against which "compliance" can be measured. Moreover, in relation to training, for example, the measurement of the quality and relevance of training provided (essentially a performance test) will be at least as important as the quantity measure suggested here. Of course, the two tests would be likely to co-exist in practice in this case – the conceptual distinctions being drawn here between the different tests being in some cases difficult to establish in practical cases.

190. Given the nature of compliance tests, it is highly unlikely that a compliance test could be identified in relation to the ***advocacy*** role of the oversight body. The advocacy role, by its nature, is one that does not conform to set procedures against which compliance can be measured. That said, oversight bodies could, in some cases, be required to report on instances of advocacy undertaken and thus provide some basic quantitative data on the extent to which this activity is being undertaken in practice. Clearly, however, the quantity of such activity will be secondary to its quality – that is, to the issue of how influential it becomes in policy development.

6.3. Performance tests

191. Chapter three defines performance tests as being tests of the quality of compliance with formal procedures or other quality standards. It is likely that performance tests may be the most fruitful means of conducting ex post evaluation of the activities of regulatory oversight units. As indicated above, the use of compliance tests will often tell us little about the effectiveness of oversight bodies involvement, while it has also been argued that the conduct of outcome tests is likely to be impeded by the difficulty of attribution of oversight body influence within the context of an outcome to which there have been many contributors. By contrast, performance tests can allow for a qualitative assessment of the specific inputs of the oversight body.

192. Within the ***advisory*** context, performance tests could include the quality of training activities undertaken to increase regulators' understanding of regulatory quality concepts and procedures. Audits of the content of such training, as well as its targeting can be undertaken by assessors with adequate expertise in the subject matter under review. An alternative approach might be to survey participants in such training directly to assess their subjective views on the quality and utility of the material presented. Reviews of the written material published by regulators could benchmark this in terms of its consistency with best practices in terms of content and presentation and could also assess its appropriateness in terms of the stage of implementation of regulatory policies.

193. An example of performance testing of this type is provided by the Dutch "visitation" scheme. This is essentially a qualitatively based assessment of the quality of legislative processes within ministries. Thus, it can be seen as a performance test of the role of central oversight bodies and other regulatory quality improvement efforts undertaken. This scheme is widely used within the Dutch administration, being applied to the legislative departments of all ministries. They are regularly reviewed in terms of criteria including their human resource management, staff training and internal organisation. Central regulatory quality oversight units such as the Ministry of Justice's legislative quality policy department are also subject to review under the visitation scheme, providing a more direct assessment of their performance as well as the indirect assessment provided by assessing the performance of their "client" groups. The process involves review of the relevant bodies by an independent committee presided over by a former secretary of state and consisting of both former senior government officials and academics. The "visitation" is preceded by a self-evaluation by the individual departments. The committee carried out two visitation rounds, in 1999 and 2001. The reports of the committee made several recommendations to further improve the organization of the legislative function within the ministries. Such recommendations have dealt with matters such as means of improving the co-operation between law drafters and policy-makers and improved human resource management. Recommendations relating to the organization or administration of individual ministries are implemented by the ministries themselves, while more general recommendations, relating to government-wide matters, are implemented by the Ministry of Justice. One such recommendation led to the establishment of a Legislative Knowledge Centre, which fosters the sharing of legislative knowledge among law drafters and aims to form a virtual "legislative community". Another measure was the foundation of an Academy for Legislation, which provides educational programmes for both aspiring and experienced law drafters. The committee concluded that its task was essentially fulfilled, following the completion of the second round of visitations in 2001. However, the Dutch Cabinet has indicated its desire to continue a system which will provide an on-going monitoring and continuous improvement of the quality of the legislative function in the ministries and is now determining the appropriate form of such a system."[20] This appears to be illustrative, in a specific context, of a key development over time in the regulatory quality agenda: the shift from seeing quality assurance in static terms toward a more dynamic view, based on constant review and policy learning.

194. Other advisory functions often performed by oversight bodies include provision of advice to Ministers as to the merits of regulatory proposals presented for Cabinet or parliamentary consideration. Assessments of such advice are necessarily more difficult where such advice is provided on a confidential basis. However, one example of this kind of assessment is evident in the United States. There, the Office of Management and Budget is required to provide annual advice to government on the costs of all regulation currently in force. These estimates are widely published and have themselves been assessed in critical articles published by independent policy advocacy groups. Such assessment can be considered as constituting externally based performance testing.

195. The ***gatekeeper*** context, provides a number of possible performance tests for oversight bodies. Given the importance of timeliness in exercising regulatory scrutiny, the timing of the oversight body's

[20] Dutch response to the OECD Questionnaire on regulatory policy instruments

first involvement in the regulatory process can function as an important indicator of the likely effectiveness of oversight body involvement. If regulators tend to contact the unit early in the process it is likely that the regulatory quality tools have achieved a substantial level of acceptance within regulatory agencies and that the oversight body has significant opportunities to affect outcomes.

196. Second, the quality of the feedback provided by the oversight body can be assessed in terms of its impact in materially improving the regulatory analysis and/or the quality of the underlying regulatory standard. This kind of test could be conducted directly, through a review of oversight agency files, involving qualitative evaluation of the feedback. Alternatively, a subjective review based on regulatory agency views of the utility of this feedback could be conducted. This latter option is likely to be less reliable as an indicator, but may also provide important additional qualitative data on perceptions of the agency and its role. This would include such issues as the appreciation of the need and benefits of consulting early with the central unit, the value-added from the involvement of the central unit and the like.

197. A third possible test is the comparison of ex ante and ex post assessments of regulatory impacts. This test has been highlighted in a previous chapter as a performance test in relation to RIA. However, it can also be seen as an important indicator of oversight body performance, at least in circumstances in which these bodies have a substantial role in assessing and approving the quality of the RIA prior to its finalisation. In such cases, the authority brought to bear by the oversight body has, as its necessary corollary, a substantial degree of responsibility for the final product. The impact of the oversight body in areas such as the provision of training and guidance publications also points to its responsibility for the accuracy of the resulting analyses.

198. Performance tests in the ***advocacy*** context would seek to measure the extent to which regulatory quality processes or initiatives proposed by the oversight body were adopted in practice by government, as well as, potentially, incorporating ex ante assessment of the degree of consistency of these proposals with established good practices and/or promising innovations in the international context. That is, they would constitute partial measures of the expected degree of improvement in regulatory policy deriving from oversight body activities in the advocacy context.

6.4. Function tests

199. Function tests are defined as those that directly measure the degree to which underlying objectives of regulatory quality policies are being achieved. As discussed above, there are substantial practical and conceptual difficulties in defining and conducting tests of this kind in relation to oversight body activities. That said, the following possibilities can be considered:

200. In relation to the ***advisory*** role, a key functional test is the extent to which initial regulatory proposals reflect an understanding of the regulatory quality principles disseminated by the regulatory oversight body as well as any direct inputs that may have been undertaken in terms of assisting regulators in the design and development of specific regulatory proposals. Where regulatory oversight bodies have the responsibility to assess RIA, review of the quality of initial RIA documents presented – i.e. prior to any changes undertaken in response to specific scrutiny and comment by the oversight body – may constitute a useful test of the effectiveness of the advisory function.

201. In relation to the ***gatekeeper*** role, one objective test would be to attempt to determine the extent to which the oversight body's review results in changes to regulators' rules. This is, in essence, a direct and positive measure of impacts on regulatory quality outcomes, and could be conducted in relation to a wide range of oversight body interventions, rather than being confined to its RIA responsibilities alone. A second form of function test would, in one sense, constitute the mirror, or negative, of this test. That is, one would seek to measure the number of times regulations that have passed through the scrutiny processes

are ultimately varied or rescinded by means that are exogenous to the regulatory policy (i.e. the regulatory quality assurance process narrowly defined). Such exogenous means would essentially comprise different types of legislative or judicial review (including quasi-judicial processes such as administrative appeals bodies). In effect, this test would be a measure of the number of times that the regulatoryquality assurance system had broken down. It would therefore be an example of a test that is subject to the difficulty of attribution, since such a breakdown reflects on the regulator and, potentially on the stakeholders that have been engaged in consultative processes, as well as the oversight body. However, to the extent that the oversight body is given a formal role as "gatekeeper" to ensure against failures at earlier stages in the process, such tests would represent a valid measure of its performance.

202. A third kind of function test associated with the gatekeeper function involves assessing quantitative data on the economic performance (i.e. net benefits or cost-effectiveness) of regulation over time. This is both a direct performance measure and one that is, at least theoretically, quantifiable. To the extent that the gatekeeper is successful in preventing ineffective regulations being made, it could be expected that the *average* cost effectiveness of regulations flowing through the system would rise.

203. There are only very few examples of evaluations of oversight units' effects on regulatory outcomes. However, one of them is consistent with this type of function test. Farrow (2000) examines the effect of regulatory reviews by the United States' regulatory oversight body, the Office of Management and Budget (OMB). The study is based on a database of 69 regulations proposed by several US agencies and reviewed by the OMB. Seven of the 69 were rejected by OMB – that is, they were sent back to the proposing agency for further consideration. Eventually all seven of these draft regulations were abandoned. The study examines several potential effects of regulatory review, notably including whether rules with poor cost-effectiveness are more likely to be rejected and whether the cost-effectiveness of rules improved during the regulatory review process.

204. The results reported by Farrow suggested that the regulatory review process had, at best, a slight impact on cost-effectiveness. Rejected rules were only slightly less cost-effective than rules that were allowed to proceed, and the cost effectiveness of rules was not found to improve during the process. However, the findings are presented with several qualifications, based on the sample size and the age of many of the rules in the database (a proportion dated from the early 1980s). Importantly, the study cannot, by definition, take account of the potential "hidden" effects of the existence of the review process on the quality of proposed rules. These effects will operate to the extent that regulators within an agency are dissuaded from proposing regulations that cannot be shown to be reasonably cost effectiveness because of their awareness of the RIA scrutiny to which they will be subjected and the risk that rules will be returned to them if they are deemed not cost-effective. Concern to avoid negative consequences such as intra-governmental sanctions or negative impacts in their relations with stakeholders (especially industry groups) could well lead them to be less willing to propos such regulations. Thus, although this study is creative and its methodology is interesting and potentially useful, with the data that are currently available, it is not conclusive. An interesting potential test of the "hidden" impact noted here would be to conduct a time-series analysis, comparing the estimated cost-effectiveness of regulations made before and after the implementation of RIA or else the adoption of substantial upgrades to an existing RIA process.

205. In relation to the ***advocacy*** role, the key functional test would be the direct identification of areas in which institutional or procedural improvements to regulatory quality had been brought about following advocacy by the regulatory oversight body. This is clearly an area in which numerous players and factors are likely to be significant in determining the outcome. However, it is likely that the regulatory oversight body will in many cases originate, or be an early advocate, of reforms and that their role in developing the concept and leading towards its adoption by government may be a clear and central one. Qualitative *ex post* assessments may well be possible in this regard, in that they may provide a reasonably reliable

overview of the importance of the regulatory body's activities in this area. The Dutch "visitation" scheme, discussed above, may represent an example of such an approach.

6.5. Conclusion

206. A fundamental conclusion of the above analysis is that the evaluation of regulatory oversight bodies is a substantially more complex exercise than is evaluation of regulatory quality tools such as RIA and consultation. This is the inevitable result of the more complex role and responsibilities of these bodies –as well as the fact that those roles vary somewhat between countries and over time. The above discussion highlights three key roles for oversight bodies – those of advisors, gatekeepers and advocates. A fully functioning evaluative process for oversight bodies must necessarily encompass a review of their performance in all of these roles. This is likely to require the application of separate tests in most cases, though there are likely to be situations where a single evaluative tool or process is able to provide data in relation to more than one function. The following table summarises the potential uses of a range of tests in relation to the different roles of oversight bodies.

Table 3. Matrix of evaluative tests: roles of regulatory oversight bodies & evaluation types

	Advisory	Gate-keeper	Advocacy
Compliance	Quantitative review of training efforts, "on request" provision of assistance to regulators, etc.	Test of proportion of RIA submitted for review; timeliness of review	Extent of advocacy role – quantitative assessment of interventions in policy process.
Performance	Analysis of improvements in policy analysis due to involvement of oversight body.	Extent to which RIA are revised as a result of the function	Assessments of quality of interventions on regulatory policy issues
Function	Analysis of improvements in regulatory proposal due to oversight body assistance in their development	Extent to which regulatory proposals change as a result of RIA assessments	Specific improvements to institutional or policy arrangements due to advocacy.

207. A general issue in relation to evaluation that is likely to be important in relation to oversight bodies is the question of the objectives against which evaluation should be carried out. The above discussion indicates that oversight bodies might be tasked with several objectives – raising the issue of how to weigh different performances against different objectives. A variant of this problem is that the oversight body may have achieved important gains in areas beyond those with which it has formally been tasked. The question must arise as to how, if at all, such benefits should be incorporated in the overall evaluation of its performance.

7. TRENDS AND CHALLENGES: TOWARDS MORE EFFECTIVE MONITORING OF REGULATORY PERFORMANCE

208. This chapter has two purposes. Firstly, it highlights trends in ex post evaluation arising from the discussion contained in previous chapters, as well as identifying challenges to the development of evaluation structures. It tries to compare what is happening at the level of individual tools and institutions and comment on whether any visible trends are complementary or antagonistic, and why.

209. Second, it focuses on the efforts of some countries to carry out more systematic monitoring of the performance of their regulatory quality systems as a whole, for example in the form of regularly reported performance assessments of all or selected regulations, embracing aspects such as rates of compliance with quality assurance processes and qualitative and quantitative assessments of actual regulatory quality. As will be seen, such systems can have broader accountability implications for governments and may form the link between centre-of-government concern regarding the implementation of regulatory policies specifically and broader issues of the quality of governance. Thus, significant additional dynamics may come into play via the attempt to adopt wider-ranging assessments of regulatory quality.

7.1. Political commitment, institutional and cultural challenges

210. Limited ex post assessment is a general characteristic of government policy and programme activity. The relative lack of ex post assessment of regulatory quality tools and institutions is thus unsurprising within this broader context. Arguably, the relatively recent adoption of many of these policies and institutional structures should lead us to expect even less evaluative activity to have been taken out to date than the average in relation to government policies and programmes more generally. Moreover, there remains a substantial challenge in terms of putting regulatory policies firmly on the broader public governance agenda. Unless policy-makers regard regulatory policy tools and institutions as fundamental to the quality of governance, they are unlikely to be persuaded to divert more substantial quantities of scarce evaluative resources toward these tools and institutions.

211. Moreover, even where their importance is recognised, the evaluation of regulatory tools and institutions generally fits badly with the policy cycle and attention span of policy-makers. Thus, a significant barrier to greater ex post evaluation activity arises from a lack of interest in, or commitment to, this activity on the part of policy-makers. In this context, it can be speculated that it may not only be an increasing recognition and acceptance of "best practices" in this area that will gradually lead to accommodations in policies, but also the impact of regulatory failures in illustrating the policy risks associated with failure to ensure that regulatory quality tools and institutions are being applied systematically, appropriately and effectively.

212. Another important barrier to more complete *ex post* analysis is institutional: the need to find a home for audits of regulatory performance, cost, and other implications of regulation. "Line" ministries responsible for preparing the rules in the first place sometimes appear reluctant to undertake such studies. They rarely have the budget to do it, and many in the ministry would think it to be beyond their mission. Also, some may question whether the promulgating ministry would have conflicts that would get in the way of a balanced assessment. On the other hand, feedback from well designed regulatory evaluations clearly has the potential to improve performance and should be of substantial interest to agencies that face appropriate incentive structures.

213. The difficulties of ensuring objectivity are also seen in relation to evaluation of regulatory quality tools. For example, in relation to regulatory impact assessment, the obvious home for review activity might be considered to be the regulatory oversight body, since it inevitably is the main source of expertise on the issue. However, such bodies are likely to face at least two separate incentives to arrive a more positive outcome than might reasonably be supported. First, they are inevitably advocates for the process, since RIA is fundamental to the regulatory policy agenda. Second, they are themselves engaged in the RIA process in almost all cases – through their advisory functions (providing training and guidance to regulators) and, in many cases, through the "gatekeeper" function, requiring them to assess and approve RIA documents.

214. Finally, "cultural barriers" is often an important constraint on *ex post* evaluation of regulatory tools and institutions. In the background survey for this report, countries noted that evaluations can face problems and resistance from target institutions who believe they are subject to "yet another reporting requirement" while regulators may also be concerned at the consequences of negative assessments of their success in applying regulatory policy tools. Embedding a positive approach toward systematic evaluation activities and evidence-based policy making in the administrative culture is necessarily a long-term challenge. However, as some of the following sections indicate, these well-known problems of policy analysis can to a considerable degree be reduced by the regulatory process itself.

Resources and methodologies

215. Resource requirements and methodological challenges also seem to play an important role. Countries consistently report resource requirements and methodological difficulties as constituting significant challenges in attempting to evaluate regulatory tools and institutions – indeed this constitutes the most widely reported concern, as indicated in Figure 4, below. In fact, these are two, quite distinct concerns in practice.

216. In relation to resource requirements, it has been argued above – as well as elsewhere in the general literature on evaluation – that resources committed to evaluation constitute investments that, if properly targeted, may yield substantial rates of return. That is, if evaluation yields policy improvements, the benefits in terms of programme cost savings and/or increased degrees of attainment of underlying objectives should more than offset the resource commitment to the evaluation process. However, there may remain substantial difficulties in having such a view accepted within the context of the budget allocation process – while the potential for evaluations to yield negative conclusions inevitably also brings with it the prospect of substantial political costs being an outcome of devoting resources to this activity.

217. Resources devoted to evaluation of regulatory quality tools must also be diverted from other uses in many cases. That is, these resources must be obtained via the competitive bidding processes of the budget system. This implies that the perceived productivity of evaluation must not only be positive but higher than that of competing bids for budget resources if they are to be allocated to this use. As well, the budgetary process is necessarily a highly political one. This means that evaluation must be perceived as being of high value within the political frame of reference if resources are to be allocated to it. There are clearly substantial disincentives in this regard, since negative evaluation results inevitably highlight past policy failures and are unlikely to be welcomed by governments if they were responsible for the implementation of those policies.

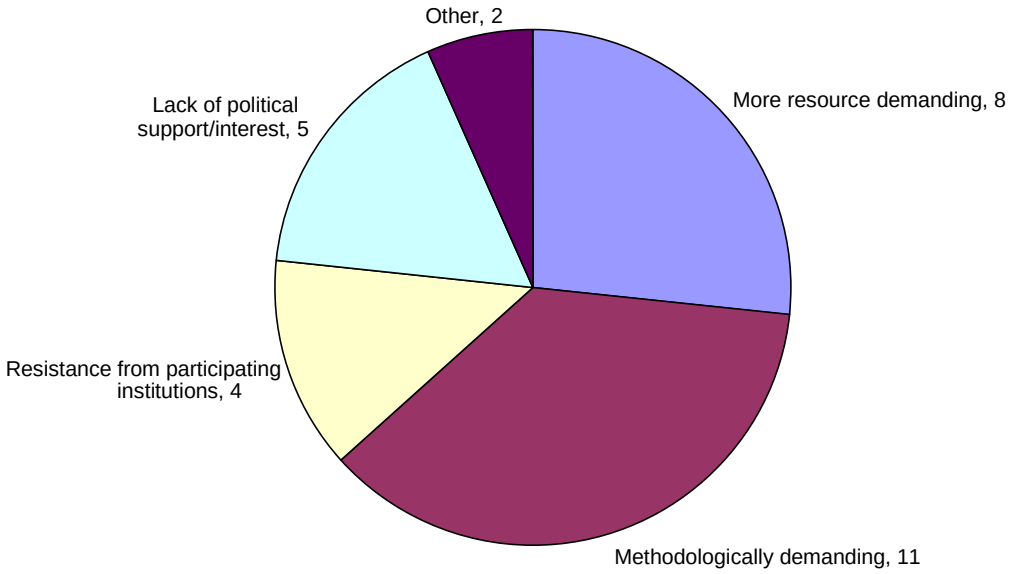
218. Several methodological challenges have been highlighted in Chapters 3 – 5. These relate particularly to the difficulties of isolating causation from correlation and in determining what objectives should be used to measure the performance of different tools. As well, it was noted that many impacts – such as that of RIA on initial policy proposals – may be unobservable in nature.

219. A related issue concerns data problems. There are often problems just finding out what the regulatory outcomes were, particularly in the context of obtaining historical data. Collecting and reporting the data consistently is yet another problem.

220. A key issue in terms of improving the basis for conducting future evaluations is the design of appropriate data collection strategies at the time of policy implementation and the adoption of strategies that will ensure that the data is collected and retained and that quality assurance is undertaken on it. That is, policy design must include consideration of the *ex post* evaluation phase if it is to be *dynamically efficient*, in the sense of allowing for an effective policy feedback loop to operate. Another data problem is the difficulty of determining an appropriate "baseline" against which to measure regulatory performance or the performance of regulatory quality tools. Determining the "counter-factual" can be an extremely difficult process, given the range of factors likely to intervene in the absence of the tool or policy being evaluated.

221. Figure 3, below, summarises the responses received to the OECD questionnaire and indicates that all of the factors discussed above can be seen as more or less widespread constraints withing OECD member countries in relation to the implementation of ex post assessment of regulatory quality tools and institutions.

Figure 3. **What have been the most significant challenges encountered in the process of assessing regulatory tools and institutions?**
(multiple answers possible)



7.2. Effective and timely monitoring of regulatory performance

222. As noted above, the development of acceptance among regulatory institutions and at the political level of the value and importance of systematic ex post evaluation is a long-term challenge involving cultural change. A distinct, but related, part of the process is the need to develop broadly applicable procedures for the routine completion of *ex post* analyses of regulatory outcomes, including the results of the application of regulatory quality tools and institutions. Responses to the OECD Questionnaire, plus

other data collected from Member countries indicate that comprehensive *ex post* analyses that examine regulatory outcomes and costs remain very rare, while equivalent analyses of regulatory quality tools are probably rarer still. Thus, the knowledge base from which to draw conclusions regarding promising approaches and practices is limited. The following surveys some of the approaches taken to date in Member countries and provides critical comment where possible.

223. Leaving aside the *ad-hoc* reviews of existing regulations carried out in most OECD Countries, it may be useful to distinguish between three approaches to reporting and monitoring regulatory performance, which vary according to their comprehensiveness and coherence as well as their integration with other evaluation and assessment programmes.

224. One approach includes ***annual government reports*** on the overall status of regulatory reform or on progress in particular programmes. Administrative simplification programmes are commonly the subject of the latter approach. In Italy for example, primary legislation introduced in July 2003 provides for the Government to submit to parliament, together with the bill for the annual regulatory simplification law, a status report outlining overall results of the simplification and regulatory reorganisation activities undertaken to date. Norway**,** following the OECD's recent regulatory reform review, is currently considering the adoption of a similar approach by which the government summarises and assesses recent regulatory reform developments in an annual report and outlines new initiatives. And in the United States, the Paperwork Reduction Act requires the Office of Management and Budget (OMB) to report annually to Congress on the paperwork burden imposed on the public by the Federal government and efforts to reduce this burden. Similarly, as noted in Chapter 3, annual reports to Congress on the aggregate costs of the regulatory structure are also required.

225. Such reporting mechanisms seem valuable in establishing a aggregate, or high-level, appraisal of progress and challenges for regulatory performance is undertaken on a regular basis. A consequence of this is likely to be that regulatory quality initiatives achieve a higher degree of prominence within government and in the wider community than might otherwise be the case. This in turn may tend to strengthen the policy over time. This kind of reporting mechanism has the disadvantage of providing limited opportunity for comprehensive reviews of the performance of individual regulations or regulatory quality tools. That is, a focus on the aggregate outcomes of regulatory activities, or of regulatory policies, may obscure rather than elucidating the performance of particular tools, programmes or institutions.

226. ***Review clauses*** constitute a second approach to more systematic monitoring of regulatory performance. Review clauses are requirements contained within regulations themselves for reviews to be conducted within a certain period. They can act as a powerful adjunct to *ex ante* RIAs by checking the performance of regulations against initial assumptions. As well, they constitute a mechanism to contribute to the dynamic efficiency of regulatory structures, by ensuring that the continued appropriateness of regulations is measured against current circumstances and new regulatory (and non-regulatory) options. Examples on this include Japan's recent regulatory reform programmes have required the inclusion in new regulations of a fixed schedule for future review. Much regulation has already incorporated requirements for *ex post* review after a fixed period of time, with review periods ranging from about 3 to 10 years after introduction. Also, the United Kingdom has recently taken steps to make systematic the use of automatic review mechanisms. RIA requirements implemented during 2000 include an obligation for regulators to set out how any proposed regulation would be monitored and reviewed. Moreover, recent policy proposals would require every government department to conduct an ex post review of the impact of major pieces of regulation within three years of their implementation. Finally Australia, the 1995 National Competition Policy agreement established a highly structured generalised review process according to which a wide range of legislation that contains restrictions on competition was reviewed and substantial reform undertaken. While this process has itself taken almost ten years to substantially complete, the NCP agreement also requires that all legislation that was subjected to the review process be subjected to further

reviews on at least a 10 yearly cycle. Also in Australia, several State governments automatically "sunset" their subordinate legislation. That is, regulations have a fixed life of between 5 and 10 years and must be formally re-made at the end of that time. Re-made regulations are treated similarly to new regulation and must pass through full RIA analysis.

227. In addition to supplementing *ex ante* assessment by checking regulatory performance against initial objectives and assumptions, these kinds of reviews may also be a suitable platform from which to evaluate the regulatory tools and institutions applied during the course of policy-making. That is, to the extent that such assessments indicate widespread problems with regulatory quality, they necessarily call into question the practical effectiveness of the regulatory quality tools that have been applied. In this respect, they can be seen as forming an important kind of "outcome test".

228. A potential third approach to monitoring regulatory performance, including the performance of regulatory quality tools, would be characterised by being more comprehensive (i.e. including all regulations) and consistent (i.e. subjecting regulations to identical review criteria), and providing a framework for assessing regulatory performance in a broader "good governance" and policy performance context. The latter would include transparency and participatory elements, and could have the potential to allow policies pursued via different policy tools (e.g. regulatory vs budgetary) to be assessed and compared.

229. The approach taken in under the ***Program Assessment Rating Tool (PART***) launched in the United States in 2002 comes close to these principles. Within PART, "Regulatory Based Programs" constitute one out of seven different types of Federal Programmes. Using PART, each resource management office (typically departmental level) of the Office of Management and Budget is obliged to evaluate all programmes under their portfolio over a five year cycle. PART comprises a number of assessment criteria on programme performance and management. Most assessment criteria are identical for all programmes, regardless the tools applied. Fundamentally, PART is focused on programme outcomes. This allows for a more comprehensive approach to assessing programme performance, for example looking at the combined effect of several regulatory measures taken in pursuit of one particular policy goal. Under PART, all Federal programmes are rated on a scale from 0-100% according to their scores on four dimensions: Program Purpose & Design; Strategic Planning; Program Management, and Program Results/Accountability. Programmes are also provided an overall rating (i.e. effective/moderately effective/ adequate/ineffective/results not demonstrated). These ratings are used to propose legislative revisions and new funding levels, as well as management or programme improvements.

230. These systematic efforts point to the fact that, to a considerable degree, the extent of the data and institutional challenges identified in the previous section, and possibly the "cultural challenges" as well, can be reduced by improved design of the regulatory process itself. As part of the regulatory process, regulators should consider the design of a potential *ex post* analysis of the regulation and the applied regulatory tools: When it should be done, how it could be done, and what it would cost. With this information in hand, policy-makers can decide at the time the regulation (or other policy measure) is issued whether to invest in data and model development that would permit an *ex post* analysis. A high quality RIA, for example, will include data on the pre-regulatory environment and models establishing a baseline. In short, the best time to begin an *ex post* evaluation of a regulation is before the regulation becomes effective.

231. These conclusions relating to regulator performance at large appear to be applicable particularly to the question of conducting evaluations of regulatory quality tools and institutions. The RIA process is, in many or most countries, characterised by a high degree of specification of objectives, procedures and methodological and other compliance requirements. The opportunity to specify ex post assessment criteria and ensure the collection of appropriate data as part of the process would seem to be substantial. Similarly,

consultation processes offer clear opportunities for follow-up with stakeholder groups of a kind that can contribute significantly to the assessment of their practical effectiveness. Incorporating such aspects in the design of these tools would clarify both government commitment to the ex post analysis process and the aims and objectives of such processes.

232. That said, the fundamental need is to be able to assess regulatory governance policies in the aggregate, rather than simply focusing on the performance of individual quality assurance tools such as RIA or consultation. This issue relates both to the inter-dependence (and perhaps the partial substitutability) of individual tools and to the fact that regulatory governance necessarily involves the assessment of the use of other non-regulatory tools, as well as the use of regulation itself. This latter perspective is clearly a corollary of the underlying concepts of the substitutability of policy instruments and the need to adopt a rational, benefit/cost based approach to the issue of choosing between different policy instruments.

8. CONCLUSIONS

233. This report has argued that knowing more about the efficiency and effectiveness of regulatory tools is increasingly important to support policy-makers in improving regulatory outcomes and reducing the risk of regulatory failures. The relevance of better, empirically based information of the performance of regulatory policy tools stems not only from the significant resources invested in regulatory management systems established in most OECD countries, but more importantly from extensions in recent years to the scope and reach of regulation as a policy instrument.

234. The purpose of the report was threefold: to provide an overview of OECD Countries' experiences with *ex post* evaluation of regulatory tools and institutions; to develop a conceptual and analytical framework to clarify the range of potential evaluation tests; and to propose guidance for government strategies to evaluate regulatory tools and institutions.

Survey results indicate growing recognition of evaluation benefits, although current activities are recent, sporadic and ad-hoc

235. Results of a survey carried out in the preparation of this report showed that, seen across OECD Countries, nearly all regulatory tools and institutions have been subject to evaluation efforts. The most commonly evaluated tools are RIAs, consultations procedures and simplification mechanisms. The most frequently evaluated institutions are independent regulators and enforcement agencies. Some evaluations are broad based, embracing different sectors and regulatory areas, while other evaluations are focussing on the application of regulatory tools or institution in one specific area. Methodologically, surveyed countries favour a non-prescriptive approach given the potentially wide ranging differences between different regulations, regulatory tools and institutions in terms of their objectives and design.

236. *Ex post* evaluation activities are recent and uneven. To the extent that specific strategies to evaluate regulatory tools and institutions exist, the strategies are part of broader regulatory and/or evaluation policies. Results of the survey showed no clear trend to favour the use of any particular evaluators. Different countries use different solutions. Responsibilities for monitoring and implementing evaluations of regulatory tools and institutions are sometimes held by specific evaluation units or agencies in some countries, with ministries of finance and/or national auditor institutions, with central regulatory oversight units or with the individual ministries. Where responsibilities for evaluation lie with the ministry in which the programme or policy being evaluated is found, the evaluation itself is frequently outsourced. By contrast, independent audit agencies which are responsible for evaluation functions often also carry out the evaluations.

237. Survey responses indicate a growing recognition of the importance of *ex post* evaluations of regulatory policies, suggesting that the level of activity is increasing and that there is a clear interest in learning from both the experiences of other countries in designing and implementing evaluations of regulatory quality tools and the theoretical writings of experts in this field.

Evaluation challenges are similar, indicating scope for cross-country learning

238. OECD Countries seem to face similar challenges in their pursuit of evaluating regulatory tools and institutions. Most notably, countries point to methodological difficulties (including data collection

strategies), institutional constraints and cultural barriers posing the most significant challenges encountered in the process of assessing regulatory tools and institutions. At the same time, the approaches taken in many evaluation projects are developed *ad hoc* with the generation of no or very little quantitative or comparable data. Thus, there seems to be significant scope for the dissemination and further development of evaluation methodologies.

The report develops an analytical framework to be used as a basis for organising and timing evaluations of regulatory policies

239. The report proposes a three-part analytical framework to assist in classifying and analysing the range of possible evaluative tools in relation to regulatory quality tools and institutions. This comprises: 1) compliance tests, which measure formal compliance with the specific standards and requirements of regulatory quality tools – such as RIA – as set out in laws or policy guidelines; 2) performance tests, which measure the quality of the analysis and related activities undertaken; and 3) outcomes tests, which measure the impact of the regulatory quality tool or institution on the regulatory outcome – i.e. its degree of success in achieving the underlying objective of improving regulatory quality and, hence, the extent to which regulation achieves its objectives.

Identification of a range of specific tests to evaluate RIA, consultation mechanisms, and regulatory overview bodies

240. By combining data taken from responses to a survey of Member country experiences in this field, more general data on evaluation practices and theoretical discussions commissioned from academic and other experts in the field, the report identifies a range of specific tests to evaluate the two most important regulatory tools – RIA and consultation mechanisms – and on one of the most important and increasingly applied regulatory institutions, the regulatory review body. An overview of the various tests are presented in Table 4 below.

241. Given the central role that RIA has assumed within the rule-making processes of most OECD Member countries, and the lengthy experience which many countries now have with the use of this tool, the relatively low level of evaluation activity apparently being undertaken is likely to constitute an important barrier to the dynamic evolution of the RIA tool. Examples cited in the report above show that significant improvements to RIA models have resulted from the conduct of such evaluations, yet they remain ad hoc and infrequent. Given this background, and the fact that regulation itself is increasingly subject to regular review requirements, there may be merit in adopting similar, regular review requirements in respect of the RIA tool itself, as a means of ensuring that options for benchmarking and improvement are provided.

242. Consultation has a substantially longer history as an element of the regulatory process than does RIA. Despite this, the application of evaluative tools to the use of consultation does not seem to be substantially further advanced than is the case with RIA. Different consultation processes have different strengths and weaknesses, and consultation can serve differing objectives. As a result, a range of consultation strategies are increasingly used in combination to achieve regulatory quality outcomes. This complexity in the nature and uses of the consultation tool implies an added degree of difficulty in identifying and carrying out appropriate evaluation. As for RIA, the report identifies a wide range of potential evaluative tools for use in relation to public consultation. A sophisticated combination of these tools that is also tailored to the specific goals underlying the use of consultation in a particular country, is likely to represent the most effective approach to the evaluative process.

243. The establishment of regulatory oversight bodies represent a relatively recent but, among OECD Countries, converging institutional response to the complexity and multiple interests involved in regulation-making. In the broadest terms, their remit is to bring a strategic and "whole of government"

perspective to bear on the processes associated with employing the regulation-making tool of government. The report concludes that the evaluation of regulatory oversight bodies is a substantially more complex exercise than the evaluation of regulatory quality tools, such as RIA, and consultation, as a result of the more complex role and responsibilities of these bodies. The report highlights three key roles for oversight bodies – those of advisors, gatekeepers and advocates – and argues that a fully functioning evaluative process for oversight bodies must encompass a review of their performance in all of these roles. This also raises the issue of how to weigh different performances against different objectives.

244. Table 4, below, summarises a range of suggested tests to evaluate RIA, consultation mechanisms and regulatory oversight units, and organises them in the three part classification of compliance, performance and function tests.

Table 4. Classification and examples of tests to evaluate regulatory tools and institutions

	Compliance tests (Process focus) Evaluates compliance with formal standards and requirements as set out in laws, policies or guidelines	**Performance tests** (Output focus) Evaluates the quality of the analysis and activities undertaken	**Function tests** (Outcome focus) Evaluates effect on quality of the regulatory outcome
Regulatory Impact Analysis (RIA)	*Benchmarking RIA contents against legislated or policy-based requirements and other guidance material - includes both analytical and process requirements.*	*Quality review of RIA components, i.e. the applied cost-benefit analysis, whether the RIA avoids egregious errors* *Transparency and clarity* *Predictability tests: Comparing actual to predicted impacts* *Model test: Quality control and accessibility of data and assumptions* *Review of internal regulatory costs* *Timing: When in regulatory process were changes made*	*Frequency of changes to initial regulatory proposals during RIA process* *Audit trails: Are options identified in the RIA process subsequently evaluated against pre-specified options?* *Differences between initial and final regulatory proposal in cost-benefit terms* *Effect of RIA on administrative/regulatory culture* *Effect of RIA on quality of initial regulatory proposals*
Consultation mechanisms	*Benchmarking of consultation procedures against legislative requirements and/or other guidance material. Are minimum standards met?l*	*Formatting (comprehension test)* *Targeting* *Response profiles* *Adequacy of reasoning* *Evidence checking (Processing of responses by regulators)* *Review of internal regulatory costs* *Stakeholder assessments of quality of consultation documents and processes*	*Audit trails: Are options identified in the consultation process subsequently evaluated against pre-specified options?* *Incidence of policy changes: Have decisions been sensitive/responsive to information provided during consultation?* *Survey of stakeholder groups about the effectiveness of consultation in improving regulatory proposals*
Regulatory Oversight Bodies	*Benchmarking of activities against formal obligations – consider advisory, gatekeeper and advocacy functions specifically.*	*Analysis of improvements in regulators' RIA due to involvement of oversight body* *Number of interventions / suggested revisions over time* *Quality review of the units' activities, including advice to cabinet, training and advisory activities, advocacy function*	*Analysis of changes in regulations due to oversight body's review* *Analysis of economic performance of regulations over time* *Survey of regulators: Better understanding of regulatory quality principles due to influence of oversight body?*

245. The report argues that the techniques that are most appropriate to evaluate regulatory tools and institutions are likely to change over time, with different evaluations being suited to the initial implementation phases of a particular quality tool than are required subsequently, when experience with its use has accumulated. Thus, compliance tests are seen as appropriate early stage evaluation tools, while the focus should increasingly shift to performance and outcome tests over time, as practices mature. Country practices, however, show no clear pattern of emphasis in terms of movement between the different types of tests over time. The inconsistent application of tests across OECD countries, plus the overall paucity of experience in this area, means there is no clear evidence in favour of one particular approach to evaluation of regulatory quality tools. A combination of approaches is likely to represent the most fruitful way forward in the medium term, with the emphasis being dependent on the specific objectives being pursued by Governments at different times

Towards more effective and systematic evaluation of regulatory policies

246. This report has drawn comparisons between the evaluation of regulatory quality tools and institutions and evaluation of government activities more broadly. Relatively little evaluation tends to be completed in both areas. Some of the reasons for this relative neglect of evaluation are also common, including the potential for negative political impacts where evaluations reveal policy failures. It is likely that moves to improve the evaluation of regulatory tools and institutions will be most effective if they occur within the context of a larger effort to expand and enhance evaluation within government more generally. Within this context, an enhancement of the scope and depth of evaluation activity is a reform that is strongly consistent with the broader governance agenda, with its focus on issues such as accountability for the use of public resources and responsiveness to public priorities. Evaluation of regulatory quality tools and institutions must itself be a core part of this move toward greater accountability and responsiveness, given the potential that these tools and institutions have for improving performance in these areas, as documented in other OECD reports on this issue.

247. Countries face a number of challenges in evaluating regulatory tools and institutions, most notably related to data collection strategies, institutional constraints, and cultural barriers. Many of these challenges can be addressed through changes to the regulatory process itself, not least by making appropriate revisions to existing RIA procedures. Somewhat paradoxically, this increased emphasis on *ex post* assessments and evidence-based policy-making incorporating feedback loops to maximise dynamic efficiency makes the *ex ante* phase of regulatory design, assessment and implementation more important. It is in this phase that behavioural targets and expected outputs and outcome must be identified if there is to be a sound basis for monitoring performance along the lines of pre-defined dimensions. Thus, a focus on ex post assessment necessarily brings with it the need for greater clarity over specific regulatory objectives and more sophisticated assessments of the likely effectiveness in practice of regulatory solutions to identified problems. Closely connected to this is the need to ensure that data collection requirements are set out *ex ante* in order to ensure that an adequate data set is available to underpin sophisticated *ex post* evaluation. [21]

[21] Taking a broader perspective, the tests and approaches to evaluation discussed in this report can generically be labelled as 'administrative approaches'. Other approaches exist, however. The most important of these alternative approaches can be labelled 'democratic evaluation' and judicial oversight. All regulatory activity in democratic societies is subject to evaluation by electorates and interest groups, whose judgements can have influence on that activity through the political process. Thus, if regulators fail to communicate and consult effectively, dissatisfaction may be made manifest in one or more of the various forums of public opinion. Similarly, regulators have to conduct their business within a given legal structure, and failures of process -- including failures in communication and consultation -- may run foul of the law. Regulatory decisions might be challenged on grounds of failure of some or another aspect of 'due process' and, in considering the challenge, the Court will be required to make an *ex post* assessment of the

Introducing a draft checklist for policy evaluation activity

248. This report concludes that experience with many aspects of evaluation of regulatory quality tools and institutions is still lacking and that substantial learning is still needed to determine the most fruitful approaches to evaluation. This conclusion relates particularly to the relative effectiveness of different kinds of tests and the means by which they are best combined in practice. However, a number of key requirements for a soundly functioning evaluation process have been identified. These are, necessarily, strongly linked to accepted "best practices" in the field of public policy and programme evaluation more generally. It is possible to summarise these requirements in the form of a draft checklist. The purpose of the checklist, set out below in Box 5, is to assist policy-makers in considering:

- the design of regulatory quality tools and institutions, focusing on ensuring that design elements that favour future evaluative efforts are included at the implementation stage;

- the design of evaluation tests, including methodology, timing and responsibility for evaluation activities;

- the design of evaluation processes, including the need for timely review, on-going evaluation activity and transparent reporting of results and responses; and

- the testing of the outcome of evaluation activity.

249. The checklist is derived from the discussion and conclusions presented in this report. Its questions are phrased in broad terms, in recognition of the importance of tailoring evaluative activity to the specific circumstances of the country in question, including the nature and objectives of the particular regulatory quality assurance system and the need to ensure consistency with broader administrative, legislative and other principles and traditions.

---

relevant process. Democratic and judicial evaluations are powerful means of ensuring that egregious failures or abuses of the regulatory tools are avoided. However, they are necessarily somewhat cumbersome and/or partial in nature and are thus less effective in ensuring that regulatory actions meet high quality standards, as distinct from avoiding major failures. This leaves considerable scope for improving evaluation by adoption of one or more of the methods outlined in this paper. However, when adopting administrative approaches, it is essential to consider their role in the context of the alternative evaluation methods that exist, and consider how the various strands of evaluation fit together.

Box 5. A Draft Checklist for Evaluation of Regulatory Quality Tools and Institutions

1. Have clear objectives been identified for all major regulatory quality tools and institutions that can form the basis of an evaluative process?

2. Have performance targets been set (either qualitatively or quantitatively) for key regulatory quality tools and institutions? If not, what criteria have been identified for evaluating the performance of these tools and institutions?

3. Is there a clear data collection strategy in relation to the implementation of regulatory quality tools, or in relation to regulatory performance more generally?

4. Is there a clear policy of evaluating major regulatory quality tools and institutions?

5. If so, does the policy allocate responsibility for evaluation activity?

6. Is the responsible body (or bodies) able to take an independent and objective view in evaluating the tools and institutions in question? Is it adequately resourced to conduct high quality reviews?

7. Are evaluation methodologies or approaches specified? Are they consistent with the objectives and priorities underlying the use of the regulatory quality tool or institution being evaluated?

8. Are multiple evaluation approaches employed? If so, are they used in a co-ordinated fashion? Are evaluations sufficiently broadly based to capture all key elements of the performance of the regulatory quality tool or institution?

9. Are specific requirements in place to ensure that evaluations are conducted in a timely fashion? Is evaluation a "one-off" activity, or are tools and institutions subject to ongoing review?

10. Is the nature of evaluation conducted consistent with the stage of implementation of the regulatory quality tool(s) or institution(s) being evaluated?

11. To whom are evaluation results reported? Is the process transparent? Are responsible Ministers or government agencies required to respond to evaluation results?

12. Is there evidence that policy "feedback" loops are operating effectively as a result of the use of evaluations in relation to regulatory policy tools and institutions?

250. The draft checklist represents an initial attempt to identify a range of key success factors for implementing *ex post* evaluation of regulatory quality tools and institutions. However, as this report has emphasised, there is limited practice to date with such evaluations. Thus, it is clear that further research is required in order to refine the elements of the checklist and, potentially, to identify important additional elements. Further developments in this area will allow for more specific guidance to be included in future versions of such a checklist. The draft provides a framework within which the design of an evaluation system can be considered, as well as a basis for further research aimed at identifying good practices for the evaluation of regulatory quality tools and institutions. Needless to say, the objective of such activities would not be evaluation for its own sake, but to operate as an instrument to improve the economic and social impact of regulations.

BIBLIOGRAPHY

Delphi Group. 2000. "Assessing the Contribution of Regulatory Impact Analysis on Decision Making and the Development of Regulations."

Farrow, Scott. 2000. "Improving Regulatory Performance: Does Executive Office Oversight Matter?" Working paper, Center for the Study and Improvement of Regulation, Carnegie Mellon University.

Farrow, Scott (2003): *Evaluating Central Regulatory Institutions*. Paper prepared for the OECD's project on *ex post* evaluation of regulatory tools and institutions. OECD, Paris.

Harrington, Winston, Richard D. Morgenstern and Peter Nelson. 2000. "On the Accuracy of Regulatory Cost Estimates." *Journal of Public Policy and Management* 19(3).

Harrington, Winston (2004): *RIA Assessment Methods.* . Paper prepared for the OECD project on *ex post* evaluation of regulatory tools and institutions. OECD, Paris.

Harrington, Winston and Richard D. Morgenstern (2003): *Evaluating Regulatory Impact Analyses*. Paper prepared for the OECD project on *ex post* evaluation of regulatory tools and institutions. OECD, Paris

McNamara, Carter (1998): *A Basic Guide to Program Evaluation*, downloaded from http://www.mapnp.org/library/evaluatn/fnl_eval.htm

Morgenstern, Richard D. 1997. *Economic Analyses at EPA: Assessing Regulatory Impact.* (Washington, D.C. Resources for the Future).

Magat, Wesley, Alan J. Krupnick and Winston Harrington. 1986. *Rules in the Making: A Statistical Analysis of Regulatory Agency Behavior.* (Washington, D.C. Resources for the Future)

OECD (1997): *Regulatory Impact Analysis. Best Practices in OECD Countries.* OECD, Paris.

OECD (1999): *Improving Evaluation Practices. Best Practice Guidelines for Evaluation and Background Paper.* OECD, Paris.

OECD (2003): *Regulatory Policies in OECD Countries: From Interventionism to Regulatory Governance.* OECD, Paris.

OECD (2003): *Framework of RIAs in OECD Countries.* Room document prepared for OECD's Working Party on Regulatory Management and Reform at its meeting 23-24 September 2004.

OECD (2003): *Regulatory Reform in Norway: Government capacities to assure high-quality regulation.* OECD, Paris. Available at http://www.oecd.org/dataoecd/26/63/2955875.pdf

OECD (2004): *Japan: Progress in Implementing Regulatory Reform.* OECD, Paris.

Forss K. (2004): Evaluation framework for information, consultation and public participation in policy making, in OECD (2004): *Measuring Success: Evaluation Tools for Information, Consultation and Public Participation in Policy-making.* OECD, Paris

NNR, Board of Swedish Industry and Commerce for Better Regulation (2003): How high is the quality of the Swedish Central Government's Regulatory Impact Analysis (RIA) in the business sector? Stockholm: NNR

Yarrow, George (2004): *Evaluation of Regulatory Communication and Consultation Mechanisms: Approaches, Methodologies and Some Guiding Principles.* Paper prepared for the OECD project on *ex post* evaluation of regulatory tools and institutions. OECD, Paris.

Radaelli, Claudio (2003): *How Context Matters: Regulatory Quality in the European Union.* Available from http://www.psa.ac.uk/cps/2004/Radaelli.pdf

ANNEX 1: THE OECD REFERENCE CHECKLIST FOR REGULATORY DECISION-MAKING

Box 6. The OECD Reference Checklist for Regulatory Decision Making

1. Is the problem correctly defined?

The problem to be solved should be precisely stated, giving evidence of its nature and magnitude, and explaining why it has arisen (identifying the incentives of affected entities).

2. Is government action justified?

Government intervention should be based on explicit evidence that government action is justified, given the nature of the problem, the likely benefits and costs of action (based on a realistic assessment of government effectiveness), and alternative mechanisms for addressing the problem.

3. Is regulation the best form of government action?

Regulators should carry out, early in the regulatory process, an informed comparison of a variety of regulatory and non-regulatory policy instruments, considering relevant issues such as costs, benefits, distributional effects and administrative requirements.

4. Is there a legal basis for regulation?

Regulatory processes should be structured so that all regulatory decisions rigorously respect the "rule of law; that is, responsibility should be explicit for ensuring that all regulations are authorised by higher level regulations and consistent with treaty obligations, and comply with relevant legal principles such as certainty, proportionality and applicable procedural requirements.

5. What is the appropriate level (or levels) of government for this action?

Regulators should choose the most appropriate level of government to take action, or if multiple levels are involved, should design effective systems of co-ordination between levels of government.

6. Do the benefits of regulation justify the costs?

Regulators should estimate the total expected costs and benefits of each regulatory proposal and of feasible alternatives, and should make the estimates available in accessible format to decision-makers. The costs of government action should be justified by its benefits before action is taken.

7. Is the distribution of effects across society transparent?

To the extent that distributive and equity values are affected by government intervention, regulators should make transparent the distribution of regulatory costs and benefits across social groups.

8. Is the regulation clear, consistent, comprehensible and accessible to users?

Regulators should assess whether rules will be understood by likely users, and to that end should take steps to ensure that the text and structure of rules are as clear as possible.

9. Have all interested parties had the opportunity to present their views?

Regulations should be developed in an open and transparent fashion, with appropriate procedures for effective and timely input from interested parties such as affected businesses and trade unions, other interest groups, or other levels of government.

10. How will compliance be achieved?

Regulators should assess the incentives and institutions through which the regulation will take effect, and should design responsive implementation strategies that make the best use of them.