



no 1 vol 2010

**Beyond »New Public Management«  
doctrine in policy impact evaluation**

Bojan RADEJ, Mojca GOLOBIČ, Majda ČERNIČ ISTENIČ



Creative commons)

Ljubljana, April 2010

Slovenian Evaluation Society

Kardeljeva ploščad 17, Ljubljana

info@sdeval.si, <http://www.sdeval.si>

---

CIP - Kataložni zapis o publikaciji  
Narodna in univerzitetna knjižnica, Ljubljana

005:35(497.4)(0.034.2)

RADEJ, Bojan

Beyond new public management doctrine in policy impact evaluation [Elektronski vir] / Bojan Radej, Mojca Golobič, Majda Černič Istenič. - El. knjiga. - Ljubljana : Slovenian Evaluation Society, 2010. - (Working paper / Slovenian Evaluation Society; 2010, no. 1)

Način dostopa (URL): <http://www.sdeval.si/Publikacije-za-komisijo-za-vrednotenje/Beyond-New-Public-Management-doctrine-in-policy-impact-evaluation.html>

ISBN 978-961-92453-3-0

1. Golobič, Mojca 2. Černič Istenič, Majda  
250449664

Bojan Radej\* Mojca Golobič\*\* Majda Černič Istenič\*\*\*

# Beyond »New Public Management« doctrine in policy impact evaluation

\* **Bojan Radej**, Independent Social Researcher, Ljubljana & Slovenian Evaluation Society, [Bojan.Radej@siol.net](mailto:Bojan.Radej@siol.net) (Correspondence address)

\*\* **Mojca Golobič**, Biotechnical faculty - University of Ljubljana & Urban institute of Republic of Slovenia; [Mojca.Golobic@uirs.si](mailto:Mojca.Golobic@uirs.si)

\*\*\* **Majda Černič Istenič**, Biotechnical faculty - University of Ljubljana & Sociomedical Institute of the Slovenian Academy of Sciences and Arts & Slovenian Evaluation Society; [MajdaCI@zrc-sazu.si](mailto:MajdaCI@zrc-sazu.si)

Proposal for citation: Radej B., M. Golobič, M. Černič Istenič. 2010. Beyond »New Public Management« doctrine in policy impact evaluation. Ljubljana: Slovenian Evaluation Society, Working paper vol. 3, no. 1 (April 2010), 28 pp, <http://www.sdeval.si/Publikacije-za-komisijo-za-vrednotenje/Beyond-New-Public-Management-doctrine-in-policy-impact-evaluation.html>

Any opinions expressed here are those of the author(s) and not those of their corresponding institutions.

Ljubljana, April 2010

**Abstract:** The European public sector has undergone major structural changes in the last three decades under the influence of what has been dubbed as »New Public Management« (NPM), mainly inspired from a number of private management models and practices, implemented in Anglo-Saxon countries. In Continental Europe, NPM has often been criticised for mimicking private management without due recognition to the complexity of public domain in its multiple scopes and multiple scales of operation. A large deal of problems linked to public sector's poor performance is not caused by the complexity of the challenge itself but by its inappropriate address with simplified private management models. Paper studies this problem on the case of impact evaluation of three main types of policy initiatives: programs, budgets and legislation. In impact evaluation not all measurable policy impacts are commensurable beyond their scope and scale as it is allowed in assessment of private sector's performance which operates on single scope (profit) and single scale (micro), but incommensurable and so complex. Recognising this, impact evaluation of public sector initiatives should be considerably modified. Paper discusses this need from methodological as well as from the normative point of view and concludes that public sector can be consistently evaluated and also managed from the aspect of its multi-scale and multi-scope complexity.

**Keywords:** New public management, Evolutionary approach, Evaluation

**JEL Code:** H11, B52, H83

### 1 Introduction

The European public sector has undergone major structural changes in the last three decades under the influence of »New Public Management« (NPM) doctrine, mainly inspired by a number of private management models and practices imported from Anglo-Saxon tradition. Doctrine is rooted in the conviction that private sector management is superior to public one. Aspiration to »reinvent government« by borrowing the best management practices found in private sector is the major driver behind NPM. The premise is that more market orientation in the public sector will lead to greater cost-efficiency of governments, with positive side effects on implementation of all »other« or »wider« social considerations (environmental, human, cohesion...). As Hayek wrote in *The Fatal Conceit* (1988, 1992),<sup>1</sup> neglect for cost-efficiency in a world of limited resources leads to neglect for others whose aspirations must remain unmet only because somebody else is allowed to waste scarce resources.

As a management philosophy NPM is applied with the aim to modernise the public sector, to debureaucratise government and to reduce red tape considerably. NPM principles have affected every aspect of policy cycle and so they also affected the approach to evaluation of public sector performance in its all three main areas, in impact evaluation of programs, legislation and budget. These three comprise the main routes for governmental intervention into the society. For instance, programming is the main policy vehicle for shaping our common future. Budget preparation concerns reallocation of 35-45% of annual GDP to provide for public goods and compensate for market failure. On the other side, poor

---

<sup>1</sup> Hayek F. 1992. *Usodna domišljavost*. Ljubljana: KRT, no. 69, 173 pp.

---

implementation of regulation is one among main drivers for emergence of wicked problems in public domain – this source of government failure is in particularly evident in the post-transition countries. As recognised by Slovenian government, for instance, adopted regulations are often in contradiction with each other or fail to regulate important issues for quality of public life.<sup>2</sup> There is an estimate for Slovenia that a loss arising from regulatory implementation gap in form of compensation payments from the state budget to victims of implementation gap has amounted on average to 1,5% of annual GDP since the beginning of nineties,<sup>3</sup> which amounts roughly to one third of the average annual growth rate of GDP in the same period of time.

But it turned out that in practice public and private sectors are essentially different. NPM has been criticised for mimicking private management without due recognition to the specificities of public domain and for being inward-looking and as such ignoring the wider implications. NPM doctrine sometimes serves as a justification for »market knows best« mantra which is employed as an excuse for inactivity of policy-makers and as a cover for their aversion for addressing deeply seated social oppositions. This became prominently displayed when NPM doctrine was transmitted beyond its original institutional Anglo-Saxon framework where it instigated a series of unexpected and unwanted side effects. As ideas, concepts, values, practices are transmitted from one cultural or political context to another they undergo a process of transformation. Thus, how a given management concept that is transferred between different management traditions is actually received and turned into new political practices – requires study.

Purpose of this paper is to investigate reasons and consequences of disappointing contribution of NPM to better public management in Continental Europe in the area of policy impact evaluation. Policy evaluation has been initially introduced as one of horizontal functions of NPM to provide neutral policy advice and to improve overall public sector's fulfilment of its complex tasks. Public managers and evaluators need to understand public sector's inherent complexity where can be no single privileged point of view, where nobody individually is able fully to understand social reality in its entirety.

In comparative perspective, private sector's rationality is relatively simple in its single operational scope (profit maximisation) and in its single scale of judgement (company or micro level). The narrow scope and scale of its rationality imposes accordingly simplified evaluative framework. Public sector on the other side is

---

<sup>2</sup> Računsko sodišče RS. 2007. Povzetek revizijskega poročila o preverjanju učinkov predlaganih predpisov, <http://www.rs-rs.si/rsrs/rsrs.nsf/I/K672E7926A481C380C1257298001F8274?openDocument>, [IX/09].

<sup>3</sup> Radej B. 2009. Ciljno usmerjen državni proračun. Slovensko društvo evalvatorjev, Delovni zvezek 5/2009, 33 pp., <http://www.sdeval.si/Publikacije-za-komisijo-za-vrednotenje/Ciljno-usmerjen-drzavni-proracun.html>, [IX/09].

---

complex in its multiple scopes (economic, social, environmental, and human – for example) and in its multiple scales of performance and evaluation: micro, meso, and macro level. Public manager of large-scale and multi-scope (LS-MS) policies (programs, legislation, budgets) is usually expected to simultaneously solve all different demands even though they may be contradictory from his/her point of view: such as to maintain high ability to manage structural or deep system change at macro level, and also exhibit high competence to serve different group interests indiscriminately (meso level) and also to deliver diversified services to each individual beneficiary accurately and efficiently (micro level). In fact, there are incommensurable viewpoints with regard to social reality in scale and in scope, so they provide us with views that are not reducible to a common denominator (Funtowicz, Ravetz, 1994).<sup>4</sup>

Gender equality is useful example for explaining social incommensurability. Gender equality is »intersectional« problem. Intersectionality is a sociological theory first coined by Kimberlé Crenshaw<sup>5</sup> in 1989 and by Patricia Hill Collins<sup>6</sup> in 1990's in their discussion on Black feminism, suggesting that and seeking to examine how various socially and culturally constructed categories of discrimination interact on multiple and often simultaneous areas, contributing to systematic gender inequality. So inequality should be studied incommensurably – on different scales (individual – collective) as well as from different »scopes« (race, religion, nationality, sexual orientation, class, or disability) that contribute to inequality.

Superimposition of commensurability assumption in social studies such as when expressing all social facts in a single-metric procedure in cost-benefit studies, violates the fundamental aspect of public sector's complexity. The problem is that in evaluation of LS-MS policies it is necessary not only to assess diverse and multiple but fragmented policy impacts (micro level) but also to translate them into summary conclusions for strategic decisions (macro level) that inform decision-makers operating at the meso level. Resulting challenge is how to provide for sufficient synthesis when society is based on incommensurability of social values and human knowledge (Kuhn, 1970; Feyerabend, 1975)?<sup>7</sup> Possibility

---

<sup>4</sup> In Martinez-Alier J., G. Munda, J. O'Neill. Weak comparability of values as a foundation for ecological economics. Elsevier, *Ecological Economics* 26(1998):277-86.

<sup>5</sup> Kimberlé C. W. 1994. Mapping the Margins: Intersectionality, Identity Politics, and Violence against Women of Color, in Fineman, M.A., R. Mykitiuk (eds.). *The Public Nature of Private Violence*. New York: Routledge, pp. 93-118.

<sup>6</sup> Hill Collins, P. Toward a New Vision: Race, Class, and Gender as Categories of Analysis and Connection. *Race, Sex & Class*, 1/1 (1993):25-45.

<sup>7</sup> Feyerabend P. 1975. *Against Method*. London, New Left Books; Kuhn T.S. 1970. *The Structure of Scientific Revolutions*. Chicago: The University of Chicago Press, [http://sciencepolicy.colorado.edu/about\\_us/meet\\_us/carl\\_mitcham/courses\\_taught/5110/classic\\_sts/structure\\_of\\_scientific\\_revolutions.pdf](http://sciencepolicy.colorado.edu/about_us/meet_us/carl_mitcham/courses_taught/5110/classic_sts/structure_of_scientific_revolutions.pdf), [VII/08].

---

of consistent synthesis in impact evaluation of LS-MS policy proposals might offer new ideas on developing complexity based public management doctrine.

Even though evaluation has been introduced to improve public sector's effectiveness in provision of public goods it seems that it became itself a factor of its poorer performance. It is concluded by the Impact Assessment Board (European Commission) that almost 80% of evaluation studies in public sector fail to produce sound policy advice in particularly at the strategic level and in subjects that cross multiple policy scopes and scales.<sup>8</sup> Inconsistent and overly simplistic policy impact evaluation not only wastes scarce public resources. It sometimes even misleads when it falsely informs policy-makers in particular at middle and at strategic level. Recent studies<sup>9</sup> have revealed that possible reason for poor policy relevance of impact evaluation could be methodological. Evaluators, quite like policy managers sometimes inadequately distinguish between complexity of social issues in scale and in scope. This frames the *working hypothesis* of paper: a large deal of problems with public sector's consistency, and in particular those concerns, linked to impact evaluation of public sector's policy-making, is not caused by the complexity of the public domain itself but by its inappropriate address of them with assumption borrowed from private management models about commensurability of social events across all scopes and scales of their evaluation.

This paper aims to improve understanding of social in/commensurability in impact evaluation of programs, legislation and budget. To achieve this "the soft approach" is needed – an approach whose logic is simultaneously mid-range and weak. Paper will inquire different approaches to policy impact evaluation from the aspect of how they take into account social incommensurability. Results will be used to inquire possibilities for more comprehensive policy impact evaluation system. Possibility of synthesis that goes beyond social incommensurability would suggest that consistent public management is possible despite its complexity.

---

<sup>8</sup> IAB. 2008. Report for the year 2008. Commission staff working document - [SEC\(2009\)55](http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=CELEX:52009SC0055:EN:NOT), <http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=CELEX:52009SC0055:EN:NOT>, [IX/09]

<sup>9</sup> <http://www.srdtools.info/>; [IX/09].

Although the official discourse around better regulation trumpets the virtues of NPM as win-win instrument, the reality is more complicated. Adoption of NPM template has not been followed by its even implementation as many countries find it difficult to apply. The dominant evaluation criterion in NPM is cost-efficiency and its logic is technocratic. Radaelli<sup>10</sup> came to the conclusion that importing management paradigms into public sector with an implicit model of technocratic rationality in mind is a common cause of disappointment and systemic failure. He thinks that a technocratic model usually turns wrong and incomplete if it is applied in different environment with the incomparable governmental traditions.

There is a range of regulatory quality and governance concerns over appropriateness of application NPM doctrine in public sectors in Continental Europe where social incommensurability plays a stronger role compared with the Anglo-Saxon tradition. After all, NPM doctrine has been challenged already by English authors such as in »The Third Way« by Anthony Giddens,<sup>11</sup> and particularly with the rise of theories associated with social complexity. NPM has very little to show for itself: British government, for example, reported that it had spent over £500 million on management consultants in a given period of time covered by the study but could only identified about £10 million in savings that could be directly attributed to their advice.<sup>12</sup> But the main problem with NPM is its anti-democratic tendency<sup>13</sup> that can be linked to systematic discrimination of multi-scope and multi-scale character of public sector's assignments.

Some authors say that NPM has peaked and is now in decline. New logic in public sector reasoning is being formed with complexity as its theoretical background since nineties.<sup>14</sup> Management theory of social complexity and the NPM doctrine have common cause but forward different strategies for coping with public management. The NPM has routinely tended to forward a coercive audit culture in public sector while complexity theory explores difficulties in governance in a shared and open manner rather than in an atmosphere of blame and sanction. Complexity is a scientific as well as a democratic approach to policy-making. With complex approach, new governance wisdom is emerging and it chimes with

---

<sup>10</sup> Radaelli C.M. 2004. How context matters: Regulatory quality in the European union. Paper prepared for the Special Issue of Journal of European Public Policy on Policy Convergence, <http://www.brad.ac.uk/acad/ssis/research/CES/>, [IX/09].

<sup>11</sup> Giddens A. 1998. *The Third Way – The Renewal of Social Democracy*. Cambridge, UK: Polity Press, 166 pp.

<sup>12</sup> Savoie D.J. What Is Wrong with the New Public Management? *Canadian Public Administration* 38/1(1995):112-21.

<sup>13</sup> Blackman T. 2001. Complexity theory and the new public management. <http://www.whb.co.uk/socialissues/tb.htm>, [IX/09].

<sup>14</sup> Cilliers, P. 1998. *Complexity and Postmodernism*. London: Routledge.



---

Habermas' »communicative rationality«<sup>15</sup> and Dryzek's »discursive democracy«,<sup>16</sup> as well as with Emirbayer's »manifesto for a relational sociology«.<sup>17</sup>

Management problems that arise with social complexity in public domain are characteristic for their fuzziness. Savoie explained the difference between private and public sector management with the dramatic comparison. In business it does not much matter if you get it wrong 10% time as long as you turn a profit at the end of the year. In public sector, it does not much matter if you get it right 90% time because the focus will be on the 10% of time you get it wrong.<sup>18</sup> And in public sector it is practically impossible for public management to get it right all the time. This is so because problems are usually ill-defined, they have political as well as purely technical aspects, they often lack a good cause-effect knowledge base, issues in the public sector are multi-faceted and difficult to pin down, they may be solved only by producing trade-offs where one solution always invokes new problems. The knowledge domain is also ill-structured in public domain as there is no one best way to solve all problems. Many public choice problems even have no answer at all. They are »wicked«<sup>19</sup> because they are deeply embedded in our societal structures, uncertain due to the hardly reducible structural uncertainty they include, difficult to manage with a variety of actors with diverse interests involved. Examples of wicked problems are the energy problem and climate change (is its cause anthropogenic or not?), labour market with increasing trend towards precarious employment forms, unsustainable patterns of mobility. Gender equality is another example: in theory of intersectionality, classical models of oppression within society do not act independently of one another. Theory holds that the classical models of oppression within society, such as those based on race/ethnicity, gender, religion, nationality, sexual orientation, class, or disability do not act independently of one another. Forms of oppression interrelate creating a system of oppression that reflects the intersection of multiple forms of discrimination. Such consistency problems in public domain are called by Ravetz »post-normal« – facts about societal matters and their cause-effect explanations are uncertain, values in dispute, stakes high and decisions urgent.<sup>20</sup>

Increasing complexity in public domain is diminishing rapidly public manager's capacity to ensure policy coherence in particular with application of simplistic

---

<sup>15</sup> Habermas J. 1979. *Communication and the Evolution of Society*. Boston: Beacon.

<sup>16</sup> Dryzek, J. S. 1990. *Discursive Democracy*. Cambridge: Cambridge University Press.

<sup>17</sup> Emirbayer M. 1997. Manifesto for a Relational Sociology. *The American Journal of Sociology*, 103/2(1997):281-317. <http://links.jstor.org/sici?sici=0002-9602%28199709%29103%3A2%3C281%3AMFARS%3E2.0.CO%3B2-A>, [IX/09].

<sup>18</sup> Savoie, Ibid.

<sup>19</sup> Rittel H., M. Webber. Dilemmas in a General Theory of Planning. *Amsterdam: Elsevier, Policy Sciences*, 4(1973):155-169.

<sup>20</sup> Ravetz, J. What is Post-Normal Science? *Futures* 31/7(1999):647-654.

---

NPM tools. Governance in public sector is growing increasingly fragmented. There is an irresolvable gap between wide mission of public sector and its actual piecemeal performance. For a number of authors efforts to reform public sector along market lines have exacerbated the problem of delivering policy coherence, democratic governance and high quality provision of public goods. Di Francesco understands these systematic difficulties as an indication of the decline of policy coherence thesis in NPM.<sup>21</sup> This induces one to inquire for new management concept that will be based on specific needs and complex rationale of the public issues.

This broad problem will be further researched on the narrower case of impact evaluation of program, legislation and budget proposals as three the main areas of policy making and government interventionism. Bentham was the first who advised the government to measure the effects of its proposals on individuals, sum these effects across all relevant individuals, counting each equally and adopt the policy proposal if the net increase in happiness is positive.<sup>22</sup> »Cumulative« strategic evaluation emerged as a distinct area of professional practice in the post-war years in North America. In NPM, policy impact evaluation is considered as a part of policy cycle consisting of programming – evaluation – implementation. Evaluation is the act of making a value judgment backed up by evidence. To evaluate is to make an explicit judgement about the worth of government proposals by collecting evidence, systematic assessment and synthesis of their worth or merit to determine if acceptable standards or evaluative criteria have been met.<sup>23</sup> Evaluation from this perspective has as its purpose to learn through systematic enquiry what works in what circumstances<sup>24</sup> and how different measures and interventions can be made more effective and how to better design, implement and deliver public programmes and policies.

However, as already noted, effective contribution of policy impact evaluation to public sector consistency is very disappointing. Deficiencies of contemporary evaluation systems can be to a large degree ascribed to aggregation problems which arise when abundant but rarely compatible empirical and qualitative evidence obtained in detailed multi-criteria assessment needs to be synthesised with the purpose of formulating summary conclusions and recommendations from

---

<sup>21</sup> Di Francesco M. Process not outcomes in New Public Management? »Policy Coherence« in Australian Government. An Australian Review of Public Affairs. 1/3(2001):103-16, <http://www.australianreview.net/journal/v1/n3/difrancesco.pdf>, [IX/09].

<sup>22</sup> Collard D. Research on Well-Being: Some Advice from Jeremy Bentham. *Sage, Philosophy of the Social Sciences* 2006; Volume 36 Number 3 September 2006 330-354

<sup>23</sup> Murari Suvedi. Introduction to Program Evaluation Department of Agricultural and Extension Education, Michigan State University East Lansing, MI 48824 <http://www.ag.ohio-state.edu/~brick/suved2.htm>

<sup>24</sup> TI, GHK, IRS. 2003. The Evaluation of Socio-Economic Development - The Guide. London: Tavistock Institute - TI, GHK, IRS, [www.evaled.info/](http://www.evaled.info/), [IV/06].

---

the aspect of public sector policies on the overall system (societal) consistency. Scriven<sup>25</sup> thinks that there is an urgent need to look carefully at the foundations of the aggregation and synthesis methodology in assessment of LS-MS policies. Many of those currently involved in the evaluation of public programmes' impact to the overall social welfare have had significant difficulties in summarizing assessed impacts into synthesized evaluative findings. Each programme addresses incommensurable viewpoints with regard to many social realities, which provide evaluators with very different »*numeraires*« and macro-views of the world which are not reducible to common denominator (Funtowicz, Ravetz, 1994).<sup>26</sup> This recalls a standard social choice problem in economics. Arrow (1951)<sup>27</sup> proved that it was impossible to scale up from all individual preference functions (micro) to produce a »public interest« function (macro) that satisfied desirable properties of an aggregation process.<sup>28</sup> By the same token, Coleman (1986)<sup>29</sup> maintains this micro-to-macro link, also referred to as social causation,<sup>30</sup> is controversial and the most poorly developed part of sociological theory. Tomer (2002)<sup>31</sup> has referred to this malfunction in economy and sociology as mainstream theories most notable failure.

The lack of explicit justification of the aggregation procedure in research of social complexity is the Achilles heel of the evaluation effort.<sup>32</sup> Neither of the opposing views (micro/macro) can avoid a reduced examination of a social complexity. A disadvantage of the macro assessment of the public policy is implicit assumption that it produces only one homogenous and easily aggregatable (commensurable) impact. For example when all policy impacts are converted into monetarised costs and benefits. A result of this is the lack of structure in social research and lack of heterogeneity due to the uniform treatment of micro-events. This is known in evaluation studies as the »macro-bias« (Elzen et al. 2002).<sup>33</sup> On the other hand, a

---

<sup>25</sup> Scriven, 1994.

<sup>26</sup> In Martinez-Alier et. al., 1998.

<sup>27</sup> Arrow K. 1951. *Social Choice and Individual Values*, (2<sup>nd</sup> ed., 1963). New Haven: Yale University Press. 124 pp. <http://cowles.econ.yale.edu/P/cm/m12-2/index.htm>, [IX/09].

<sup>28</sup> Evans T.P., E. Ostrom, C. Gibson. Scaling issues with social data in integrated assessment modelling. Swets & Zeitlinger, *Integrated Assessment*, 3/2-3(2002):135-50, [http://journals.sfu.ca/int\\_assess/index.php/iaj/article/viewFile/29/17](http://journals.sfu.ca/int_assess/index.php/iaj/article/viewFile/29/17), [IX/09]

<sup>29</sup> Åberg Y. Individual Social Action and Macro Level Dynamics: A Formal Theoretical Model. *Sage, Acta Sociologica*, 43/3(2000):193-205.

<sup>30</sup> Sawyer R.K. *Artificial Societies: Multiagent Systems and the Micro-Macro Link in Sociological Theory*. *Sage Sociological Methods Research*, 31/3(2003):325-63.

<sup>31</sup> In Svendsen G.T. G.L.H Svendsen, ed. 2008. *Handbook of Social Capital: The Troika of Sociology, Political Science and Economics*. Cheltenham: Edward Elgar, [http://www.ebookee.com/Handbook-of-Social-Capital-The-Troika-of-Sociology-Political-Science-and-Economics\\_351573.html](http://www.ebookee.com/Handbook-of-Social-Capital-The-Troika-of-Sociology-Political-Science-and-Economics_351573.html) [IX/09].

<sup>32</sup> Scriven M. The Final Synthesis. *Sage, American Journal of Evaluation*, 15/3(1994):367-382.

<sup>33</sup> In Schenk N.J. 2006. *Modelling energy systems: a methodological exploration of integrated resource management*. Groningen: University of Groningen, PhD Dissertation, Chapter 6, p.

---

disadvantage of the micro or bottom-up assessment of the impacts on individual project (and criteria) level is that its conclusions are based on extrapolations from non-representative individually observed cases. As a result of uncertainty at the micro level, bottom-up evaluation tends to widely »over-forecast« or »under-forecast« at the top-level (Kahn, 1998);<sup>34</sup> therefore, it is also unable to assess changes in the whole social system.

Effective example of aggregation problem in contemporary social research is again brought forward with the feminist theory of intersectionality. Intersectional theory requires to study gender inequality so that it examines how various socially and culturally constructed categories of discrimination interact on multiple and often simultaneous areas, contributing to systematic inequality. Intersectionality specifically constitutes a critical alternative to additive arithmetical frameworks (such as commensurability) involving multiple jeopardy. For instance, theory dismisses the additive claim that black women are twice as badly off than white women due to both sexism and racism. According to Prins,<sup>35</sup> intersectionality emphasizes that »the complexity of processes of individual identification and social inequality cannot be captured by such arithmetical frameworks«.

Formal reason for aggregation problem in research of social complexity is that causality in public affairs is not linear as it is in private sector but non-linear. Non-linearity refers to situations in which: (i) qualitatively diverse causes contribute to the same effect – such as in the case of abandonment of agricultural land which is sometimes result of strictly protectionist nature conservation policy as well as autonomous result of economic factors causing depopulation; (ii) one policy (social cause) induces qualitatively different effects, such as primary or targeted impacts and also unintended or secondary impacts.

The following sections of this paper will inquire if problematic methodology of aggregation in impact evaluation can explain the considerable gap between high theoretical aspirations in evaluation and its poor practical contribution to consistency of public management. If evaluation problems in different evaluation approaches are found below as systematic, a need will arise for meta-evaluation of presently prevailing approaches to impact evaluation of public policies that is compatible with complex nature of policy challenge.

## 2.1 Budget Performance Monitoring

---

97-115, [http://dissertations.ub.rug.nl/FILES/faculties/science/2006/n.j.schenk/06\\_c6.pdf](http://dissertations.ub.rug.nl/FILES/faculties/science/2006/n.j.schenk/06_c6.pdf), [VIII/07].

<sup>34</sup> In Schenk, 2006.

<sup>35</sup> Prins, B. Narrative Accounts of Origins: A Blind Spot in the Intersectional Approach? *European Journal of Women's Studies* 13/3(2006):277-290.

---

Idea of performance based budgeting is integral part of NPM paradigm and has been first formulated in the Anglo-Saxon world in late forties. Today, all OECD countries practice one or the other variant, more or less comprehensive, of performance based budgeting. Robinson (2007) gives the following definitions for performance based budgeting. In the broad case, »it refers to public sector funding mechanisms and processes designed to strengthen the linkage between funding and results (outputs and outcomes), through the systemic use of formal performance information, with the objectives of improving the allocation and technical efficiency of public expenditure.«<sup>36</sup> He further characterizes performance based budgeting as »consist[ing] of classifying government transactions into functions and programmes in relation to the government's policy goals and objectives; establishing performance indicators for each programme or activity; and measuring the costs of these activities and the outputs delivered«.

Performance measures are most strategically useful for public managers when they help to determine how should be available public funds allocated to the various public purposes and aims. Other aims of performance oriented budgeting are to:<sup>37</sup> (i) provide public managers with capability to monitor implementation of public policies, identify potential problems, and take timely corrective action; (ii) encourage long-term thinking, make more informed decisions on policy priorities and resource allocations; help improving strategic choices; (iii) convert accountability for spending money to accountability for achieving results and direct resources to activities that meet or exceed performance standards and objectives; (iv) translate choices about goals and priorities into actions/performance objectives and communicate them more effectively to the implementation managers; when they know the basis on which they will be assessed, they are more likely to perform; (v) remove needless constraints in implementation on managers' uses of public resources to encourage management innovation and provide positive incentives to cut wasteful spending; (vi) build trust and enhance credibility of the government as a whole with taxpayers.

While it is tempting to press forward to adopt a fully fledged performance based budgeting framework in all EU member countries, there are evident risks in the move. Such a change in orientation is only possible once managers in public sector have developed a comprehensive system of performance measurement. The latter is usually lacking in EU member states either because of weaker institutional capacities, in particular in programming phase of policy cycle (new

---

<sup>36</sup> Robinson M. Performance Budgeting Models and Mechanisms, in M. E. Robinson, & M. Robinson (ed.), 2007, New York: International Monetary Fund, pp. 1-21.

<sup>37</sup> National Performance Review. Mission Driven, Results Oriented Budgeting, September 1993; CBO. 1993. Using performance measures in the federal budget process. Congress Of The United States, Congressional Budget Office, 64 pp., <http://www.cbo.gov/doc.cfm?index=10349>, [IX/09]; Diamond J. Establishing a Performance Management Framework for Government. Instituto de Estudios Fiscales, Presupuesto y Gasto Público, Working Paper 40(2005):159-83.

---

member states) or also because of different institutional context (less technocratic, more formalistic, or more consensual; see Radaelli, 2004) compared to Anglo-Saxon countries. As reported by those OECD countries with the longest history in performance monitoring, its implementation is subject to severe difficulties. In its monitoring domain, which is probably the most demanding part of performance monitoring, the methodological problems are regularly reported:

1. The ability to measure performance is inexorably related to a statement of what the agency or program is trying to accomplish.<sup>38</sup> The task of clarifying those goals is much more difficult and inherently different for public-sector agencies than for private corporations. In the private sector, the primary measure of performance for an organization as a whole is a profit. Public agencies have no such simple measure of performance. In public sector, performance must be judged against all various and equally important purposes and goals of a public policy and these are many times contrary to each other. These goals differ in intention and in effect from policy to policy, and even within a given policy there may be disagreement about the precise nature of a particular policy scope. Because of complexity it is more difficult, therefore, for public agencies to determine with any certainty what they are trying to accomplish and therefore what should be precisely achieved with the performance monitoring. This considerably diminishes possibility for successful transfer of good management practices from private to public sector in particular when this is attempted with the ignorance of complex nature of collective choice problem.
2. Setting objectives is generally only the first step in performance monitoring. Once agencies have determined what they should be accomplishing, it is perhaps even more challenging to measure progress toward those goals.<sup>39</sup> Although many public agencies collect a great deal of data, these data have typically focused on the activities of the agency rather than its results or broader societal consequences. Input, output and outcome measures in micro perspective are typically used by agencies, even though impact as macro measures would be more meaningful in system-wide evaluations:<sup>40</sup>
  - a) Inputs are describing resources consumed by the publicly financed purpose (program, organisation). They are usually commensurable in their nature, so they are easily measured, usually in terms of money. They are necessary for the achievement of objectives, but the question of how many inputs are required usually goes unanswered.

---

<sup>38</sup> CBO. 1993. Using performance measures in the federal budget process. Congress of the United States, Congressional Budget Office, 64 pp., <http://www.cbo.gov/doc.cfm?index=10349>, [IX/09]; Diamond, 2005.

<sup>39</sup> CBO, 1993.

<sup>40</sup> Diamond, 2005.

- 
- b) Outputs are immediate results of an agency's activities. Unlike inputs, outputs are often impossible to translate into money because there are no markets for most government activities. These performance indicators remain silent about the effect of agency's output on the targeted social issues.
  - c) Outcomes measurement concern the extent to which the activities (outputs) of the public agency have an intended effect for the beneficiary. That is, they focus not only on the work performed, but also on the results of that work. This requires introduction of non-governmental assessment criteria for evaluation of governmental undertakings. Still, even this kind of indicator is mute when it comes to question how well public agency's operation fits into overall social situation.
  - d) Impact performance measures are needed for identification of wide-system impacts of public program on overall welfare, system cohesion, and adaptability to external shocks etc. The problem is that these impacts are long-term and mostly indirect or they are impossible to measure and specify with certainty.
3. Multi-criteria assessment of public programs is needed but this rises problem of synthesis of the performance results not only to report to responsible implementation managers (micro view) but also to inform decisions about the budget allocations (macro view).
  4. Performance systems seem to work best on micro (organisation, project) level where there is direct accountability or clear cause-and-effect relationships between what the public agency does and what is achieved. The macro level achievements (cohesion, sustainability) are far less satisfactory covered with performance monitoring. It is in particular difficult for public agencies to link their performance results (obtained at micro level) with the budget allocation (decided at macro level) »in any meaningful way« because the relationship is not straightforward:
    - a) Poor program implementation results may be caused by the difficulty of the public issue being addressed rather than by inadequacies in the design or in policy implementation.<sup>41</sup> In such a case poorly performing policy ought to get more resources not less as monitoring results would imply.
    - b) Good performance monitoring results of a given public program shall not by itself grant an access to public funding. Results are usually exhibited against criteria that are selected by the implementation agencies themselves so these results can hardly be seen as »neutral« from wider aspect of view. Even when this is not the case, performance measures for

---

<sup>41</sup> CBO, 1993.



---

one budgetary item or public program are usually observed separated from other budgetary items – and so their eventual successful realisation by itself does not inevitably guarantee overall positive social impact.

Only limited macro relevance of performance monitoring is obviously linked with inconvertibility of micro level observations directly into meanings at macro level. Because of scale incommensurability, macro conclusions are not directly obtainable from micro observations but they emerge from them in more subtle way.

Conclusion is that performance monitoring poorly distinguishes between scope and scale complexity of public domain. This directs public manager as well as performance monitor to number of research questions that need study to explain how multiple-scale and multiple-scope view could be consistently implemented in performance based budgeting.

## 2.2 Regulatory Impact Assessment

Regulatory impact assessment (RIA) is an integral part of NPM doctrine. RIA has been introduced with NPM as a guide for better regulation and a decision tool in public sector. It is a method of systematically and consistently examining selected potential impacts arising from regulatory action.<sup>42</sup> Its role is to provide a detailed and systematic appraisal of the potential impacts of a new regulation in order to assess whether the regulation is likely not only to achieve direct (targeted) and prioritised objectives (lower administrative burden and costs of regulation), but also instigate »other« un/favourable effects of regulation on the overall society in terms of its stability, cohesion, sustainability. RIA thus broadens the mission of regulation from highly-focused problem-solving concerning only specified legal questions to balanced approach to regulation of the social.

In practice, however there is considerable evidence of poor implementation of RIA.<sup>43</sup> OECD (1997, 2005) presented studies that converge to a conclusion about more than a few difficulties at implementation of RIA. A point that the concept of RIA spreads from its Anglo-Saxon origins to Continental Europe is often neglected by the (so far hegemonic) »one-size-fits-all« approach which is disregarding context sensitivity and the wider implications of the public sector policy for social cohesion and. Sometimes imported RIA practices are expected somehow automatically to accommodate significantly different regulatory traditions. This is in particular evident in post-transition countries. Policy-makers

---

<sup>42</sup> Jacobs S.H. An overview of regulatory impact analysis in OECD countries, in OECD. 1997. Regulatory impact analysis: best practices in OECD countries. Paris: OECD, p. 13-30

<sup>43</sup> Rodrigo D. 2005. Regulatory Impact Analysis in OECD Countries: Challenges for Developing Countries. OECD, South Asian-Third High Level Investment Roundtable, Dhaka, Bangladesh, 33 pp.



---

who have tried simply to import RIA from different institutional contexts have found it difficult to scratch below the surface of new public management rhetoric; they found that RIA often failed to support implementation of more successful legislation and regulatory reforms.<sup>44</sup> As a result, hasty adoption of »good practices« is soon translated into implementation problems. Radaelli et al. have in particular identified Bulgaria, Lithuania, Greece, Romania, Portugal, Slovenia, and Spain as the laggards in implementation of regulatory impact assessment.<sup>45</sup> Slovenian Audit Court (SAC) has noticed poor effectiveness of regulatory impact assessment in Slovenia<sup>46</sup> because it covers too narrow scope (only what is primary targeted with legislation) and because of ignoring the wider (macro) social implications of the legislation. It pointed out illogicality that at the government level, 70% of legislative proposals are said to have, beside impacts on the targeted area, no other (secondary) impact on wider society and environment; remaining 30% are considered to have only neutral impact. Do we really need, SAC asked, such legislation that makes no difference?

Rodrigo (2005) states that the appropriate path to regulatory reform for better governance in every country depends on the consistency between its political, cultural and social characteristics »that specify a backstage for reform venture«.<sup>47</sup> In this regard, it is important to stress that there is no correct model for RIA. Two RIA institutional traditions can be distinguished:

1. Anglo-Saxon tradition, which is technocratic and derived from private sector mentality; in this tradition the overall aim of RIA is assessment of the potential economic, financial and administrative »burdens« of regulatory proposals. In this tradition, the main scope of RIA is to assist governments in making their policies more cost-efficient. This is entirely in line with NPM paradigm emphasising deregulation and degovernmentalisation of public life.<sup>48</sup>
2. In Continental European RIA tradition deregulation has in a large extent disappeared from the agenda of regulatory reform.<sup>49</sup> Not less but qualitatively

---

<sup>44</sup> Radaelli, 2004.

<sup>45</sup> Radaelli C.M, F. De Francesco, V.E. Troeger. 2008. The Implementation of Regulatory Impact Assessment in Europe. ENBR workshop, Exeter: University of Exeter, 27-28 March, 27 pp., <http://centres.exeter.ac.uk/ceg/research/riacp/documents/ImplementationofRIAENRworkshop.pdf>, [IX/09].

<sup>46</sup> Računsko sodišče, 2007.

<sup>47</sup> Rodrigo, 2005.

<sup>48</sup> OECD. 1997. Regulatory impact analysis: best practices in OECD countries. Paris: OECD, p. 13-30.

<sup>49</sup> Radaelli, 2004.

---

better regulation is promoted,<sup>50</sup> not efficiency but system cohesiveness (social, regional, territorial) is its main concern. Continental tradition<sup>51</sup> is diversified and promotes multitude of approaches in national domain: Sweden, Denmark, Netherlands, France, Germany... Here RIA obtained a new role in enhancing the ability of policy-makers to serve diffuse interests, rather than responding to narrower and more focused ones.<sup>52</sup> Efficiency is only one among relevant assessment scopes. The example from the field of gender equality: both sexes have different societal roles so regulation impacts them differently; legislation can lead to a de facto different situation in rights between women and men;<sup>53</sup> RIA should therefore broaden its scope and differentiate legislative impacts also by gender.

Different RIA templates exist and they are diversified in assessment scales and scopes. Relatively simple techniques with the narrowest scope are applied in assessment of »financial burden« of new regulation on public budget. This part of RIA intersects with budget performance monitoring, explained earlier. Another RIA template assesses only »administrative burden« that is imposed on population and organisations by a given piece of regulation. One of commonly applied methods of administrative burden assessment is based on measuring »information cost« that is a result of additional information demand (paperwork) that will be imposed on people and businesses with the adoption of regulation. Such narrow scope in RIA elevates suspicions. It is not clear why administrative burden of regulation should be chosen as a dominant assessment scope? In this way RIA ignores »other impacts« such as on opportunities of affected population and organisations. Regulation is constraining activity so it would have sense to assess its impact primarily on the opportunities of those affected. Also, why administrative burdens of a new piece of regulation depend solely on information cost? It is true that regulations in contemporary information society increasingly create and manage information flows. But it is not advisable to generalise and claim that regulation is essentially information device since information cost of regulation may be irrelevant in wider evaluation context.

RIA's scope usually does not cover »other« (non-financial and non-administrative) impacts with wider implications on society, environment; gender equality, because they are perceived as secondary or unintended and too hard to assess.

---

<sup>50</sup> Hopkins T.D. Alternative approaches to regulatory analysis: designs from seven OECD countries, in OECD. 1997. Regulatory impact analysis: best practices in OECD countries. Paris: OECD, p.123-41.

<sup>51</sup> Radaelli C.M. Desperately Seeking Regulatory Impact Assessments: Diary of a Reflective Researcher. Sage, Evaluation 2009; 15/1(2009):31-48.

<sup>52</sup> Deighton-Smith R. Regulatory impact analysis: Best practices in OECD Countries, in OECD. 1997. Regulatory impact analysis: best practices in OECD countries. Paris: OECD, p. 211-41, <http://www.oecd.org/dataoecd/21/59/35258828.pdf>, [IX/09].

<sup>53</sup> Gender in EU funded research, Cordis (<http://www.yelowwindow.com/genderinresearch>), [I/10]

---

This also means that RIA is not assessing the synergies behind a regulatory choice.<sup>54</sup> These drawbacks are particularly problematic when observed in Continental regulatory tradition and even more in post-transition countries with their poorer legislative implementation effectiveness, where »other« impacts of legislative acts of government might even have dominant role because of more abundant unwanted effects. In this respect, RIA triggers basically the same set of scope and scale related methodological problems as those presented below for SIA, so in this regard broadly the same set of research challenges is accompanied with them.

### 2.3 Strategic Impact Assessment

Associated with previously discussed meta-theoretical concerns of social complexity, there is an apparent paradigm crisis in the strategic impact evaluation (SIA)<sup>55</sup> of LS-MS programs such as in evaluation of program's sustainability, impact on social cohesion, or quality of life etc. SIA should take into account all insurmountable public scopes indiscriminately which immediately invokes an aggregation problem. It surfaces from a disagreement over assumptions about the aggregation of numerous policy impacts (micro level) into macro evaluative conclusion that should inform decision-makers who operate at meso level.

This problem can be elaborated on the case of the standard matricial impact assessment approach introduced by Luna Leopold (et al, 1971).<sup>56</sup> In his approach, assessment of program impacts involves only two scopes (economic, nature) and only one scale (micro – impact of measure  $x_i$  on nature evaluation criteria  $y_j$ ). This method introduced a detailed (micro) SIA, which presents impacts of numerous policy measures (in his case 100) onto numerous environmental assessment criteria (in his case 88). His impact matrix reaches thousands of cells with detailed estimates of a given program impacts. Nevertheless, Leopold explicitly rejected the summation of the assessed impacts into aggregate indicator of overall or system-wide program impact because he properly recognised the fact that impacts can be evaluated either with economic or environmental values (weights) which are incommensurable between each other. After more than 30 years, EU Impact

---

<sup>54</sup> Radaelli, 2004.

<sup>55</sup> Virtanen P., P. Uusikylä. Exploring the Missing Links between Cause and Effect. A Conceptual Framework for Understanding Micro–Macro Conversions in Programme. Sage, Evaluation, 10/1(2004):77–91; Hertin J., A. Jordan, M. Nilsson, B. Nykvist, D. Russel, J. Turnpenny. 2007. The practice of policy assessment in Europe: An institutional and political analysis. EU/FP6 Project MATISSE Working Paper 6, 52 pp., [http://www.matisse-project.net/projectcomm/uploads/tx\\_article/Working\\_Paper\\_6.pdf](http://www.matisse-project.net/projectcomm/uploads/tx_article/Working_Paper_6.pdf), [IX/09]

<sup>56</sup> Leopold L.B., F.E. Clarke, B.B. Hanshaw, J.R. Balsley. 1971. A procedure for evaluating environmental impact. Washington: Geological Survey Circular 645, 13 pp., [http://eps.berkeley.edu/people/lunaleopold/\(118\)%20A%20Procedure%20for%20Evaluating%20Environmental%20Impact.pdf](http://eps.berkeley.edu/people/lunaleopold/(118)%20A%20Procedure%20for%20Evaluating%20Environmental%20Impact.pdf), [IX/09].

---

Assessment Guideline still explicitly follows the same logic.<sup>57</sup>

However, rejection of aggregation of fragmented assessment results in SIA is highly problematic. Very detailed assessment of LS-MS public programme generates »information overload« and produces banal answers to complex and multidimensional societal problems.<sup>58</sup> Recently Ekins and Medhurst (2003, 2006)<sup>59</sup> proposed aggregated variant of SIA that takes into account the multiple-scope perspective on the impact side of evaluation matrix (Leopold-Ekins-Medhurst approach – LEM).<sup>60</sup> Their research was aimed at developing a tool for synthetic evaluation of structural fund expenditure's impact on regional sustainability. They derived their proposal from previous research on a strategic environmental assessment (SEA).<sup>61</sup> SEA has been accepted as a standard procedure for binary scope (cause and effect) evaluation of large scale economic policy projects on nature. They extended SEA to cover multiple public scopes indiscriminately and so they included also social and human scope in the evaluation. Results of their work have been taken into account in preparation of the guide for the evaluation of socio-economic development,<sup>62</sup> which become one of standard referred sources in impact assessment standards in the EC. However, broadening evaluation of scope from two to many scopes imposes certain methodological problems that have been overlooked by Ekins and Medhurst.

They proposed an assessment matrix that is a vertically and horizontally compacted version of Leopold matrix. LEM's criteria columns are condensed on dozen indicators that represent the four main evaluation scopes; they further allowed for vertical aggregation of assessed programme impacts. Yet, many authors made it evident that the impacts of different sectoral public policies are not homogenous<sup>63</sup> and produce differentiated impacts in scope. In multi-criteria assessment, linear relationship between cause (policy) and effect (impact) is broken. It has been elaborated theoretically<sup>64</sup> and confirmed empirically that inappropriate synthesis of evaluation results in assessment of LS-MS programmes produce different and sometimes even wrong advice to public managers.<sup>65</sup>

---

<sup>57</sup> Impact Assessment Guidelines, SEC(2005)791, March 2006 update, pp. 39-40

<sup>58</sup> Virtanen, Uusikylä, 2004.

<sup>59</sup> Ekins, 1992; Ekins P., Medhurst J., 2003. Evaluating the Contribution of the European Structural Funds to Sustainable Development. Presented at the 5<sup>th</sup> European Conference on Evaluation of Structural Funds, Budapest, June 26-27, 48 pp, [http://europa.eu.int/comm/regional\\_policy/sources/docgener/evaluation/rado\\_en.htm](http://europa.eu.int/comm/regional_policy/sources/docgener/evaluation/rado_en.htm), [IX/09]

<sup>60</sup> <http://www.srdtools.info/>; [IX/09].

<sup>61</sup> Sadler B., R. Verheem. 1996. Strategic Environmental Assessment: Status, Challenges and Future Directions The Hague: Ministry of Housing, Spatial Planning and the Environment.

<sup>62</sup> TI, GHK, IRS. 2003.

<sup>63</sup> Schnellenbach J. The Dahrendorf hypothesis and its implications for (the theory of) economic policy-making. Cambridge Journal of Economics, 29/6(2005):997-1009.

<sup>64</sup> Scriven, 1994.

<sup>65</sup> Radej, 2008.

---

Recognition of sectoral bias of policy impacts is of crucial importance because it demands evaluator of public programs to go beyond performance monitoring logic which aims to assess only achievements targeted by the program, either outputs or outcomes. Ekins and Medhurst overlooked that for example each economic measure does not only directly and predictably impact its primarily targeted goals (economic), but also causes unintended or secondary impacts when usually unpredictably affecting several »other« areas that fall under jurisdiction of other parts of the program or other sectoral policy.<sup>66</sup> As a matter of principle, institutional interventions should always be addressed in terms of their inadequacy due to their specialization against the general interest they serve (Donzelot, 1991).<sup>67</sup> Empirical studies plainly confirm the sectoral impact bias even for those policies that had previously been taken as the most neutral in scope, such as monetary<sup>68</sup> and tax policy.<sup>69</sup>

The consequence of identified public interventions' sectoral bias in scope of their impacts is that LEM's vertical summation is inappropriate. Vertical summation does not »preserve the negation«<sup>70</sup> between evaluation scopes which results in discriminated evaluation of incommensurable categories. Ekins and Medhurst have overlooked that the impacts of sectoral policies on a given assessment scope are not fully comparable and thus not strongly commensurable; part of the sectoral incomparability of impacts is due to incommensurability and these differences are structural or deep and must be preserved in a synthesis of evaluation results so they themselves can become an object of evaluation. For example, economic and social policy's impacts on the environment are not commensurable so they need to be aggregated separately (such as economic impacts on the nature separately from social impacts on the nature).

Policy impacts as incommensurable can only be comparable in an overall picture without recourse to a single value<sup>71</sup> applied to them. The difficulty with LEM is that it comprehends multi-scope aspect of programme only on the impact side of the programme and not on the causal side of policy scopes. In LEM the sectoral scopes or sectoral intentions are considered as homogenous. This is of course not

---

<sup>66</sup> Rotmans J. Tools for Integrated Sustainability Assessment: A two-track approach. Vancouver: University of British Columbia, The Integrated Assessment Journal, 6/4(2006):35–57, [http://journals.sfu.ca/int\\_assess/index.php/iaj/article/view/250/219](http://journals.sfu.ca/int_assess/index.php/iaj/article/view/250/219), [IX/09]

<sup>67</sup> In Burchell, *ibid.*

<sup>68</sup> Lucas R.E. Jr. 1972. Expectations and the neutrality of money. Blackwell, Journal of Economic Theory, 4/2(1972):103-24.

<sup>69</sup> Leith C., L. von Thadden. 2006. Monetary and fiscal policy interactions in a new Keynesian model with capital accumulation and non-Ricardian consumers. Working Paper Series No 649, 42 pp, [http://ssrn.com/abstract\\_id=908620](http://ssrn.com/abstract_id=908620), [IX/09].

<sup>70</sup> Ostmann A. 2006. The aggregate and the representation of its parts. Bonn: Max Planck Institute for Research on Collective Goods, Preprints of the Max Planck Institute for Research on Collective Goods 2007/11b, 38 pp., <http://ssrn.com/abstract=1024681>, [IX/09].

<sup>71</sup> Martinez-Alier et al, 1998.

---

the case. Sectoral programme involves incommensurability in scope already in its primary intentions (causes), not only in its primary and secondary impacts (effects) as LEM assumes.

The conclusion is that both micro approach (fragmented Leopold's) and macro (aggregated LEM's) approach to SIA fail to provide decision-maker with strategic policy advice. If detailed impact evaluation results are not summarised as in Leopold's matrix, the assessment produces findings that are too fragmented, in this way wasting insight into the structural relations between scope domains. In contrast, full aggregation, such as in LEM, causes evaluative findings that are amassed too much, wasting information on the synergy or oppositions between autonomous public scope domains. Both micro and macro type of SIA waste essential information on the scope and scale complexity of the policy program being evaluated which exposes evaluation results to manipulation or at least to deceitful interpretation. Without explanation of how different scopes of complex public programme impact each other and if they do or do not work together, it is impossible to substantiate evaluation findings nor to say anything about the overall impact of the programme proposal in wider social context.

The methodological problem is of course not due to different detailed expert assessment of impacts but arises entirely from the decision on how to aggregate fragmented results and interpret synthesis in the evaluation of complex public issues. Majority of standard SIA approaches are found to mismanage incommensurable differences of the assessed programs, such as EU's SEA (2001/42/EC), Impact Assessment Guidelines (SEC(2005)791), the territorial impact assessment (TIA)<sup>72</sup>, and ex-ante assessment of the contribution of the EU structural funds to regional sustainability.<sup>73</sup> The conclusion is that SIA is needed which will be able to cope with social complexity in scope and scale indiscriminately.

### 3. Policy impact evaluation for social complexity

Previous sections exemplified on several cases that new public management doctrine is too narrow base for developing sound policy impact evaluation approach in European Continental context because it is not appropriately reflecting complex nature of public domain. Three reviewed evaluation tools, PBB, RIA and SIA, are accompanied with similar difficulties in at least three aspects, that can be linked to public sector's complexity: (i) there is inconsistency

---

<sup>72</sup> ESPON - 3.2, 2006, [http://www.espon.eu/mmp/online/website/content/projects/260/716/index\\_EN.html](http://www.espon.eu/mmp/online/website/content/projects/260/716/index_EN.html), [IX/09]

<sup>73</sup> GHK, PSI, IEEP, CE, National Evaluators. 2002. The Contribution of the Structural Funds to Sustainable Development: A Synthesis Report to DG Regio, EC. 2002. Volume 1-2. London, Brussels, [http://ec.europa.eu/regional\\_policy/sources/docgener/evaluation/doc/sustainable\\_annexes.pdf](http://ec.europa.eu/regional_policy/sources/docgener/evaluation/doc/sustainable_annexes.pdf), [IX/09].

---

between evaluation philosophy and institutional characteristics in which public choice takes place; (ii) there is also scoping problem in evaluation in particular regarding coverage of unintended, secondary, wider and long term social impacts in evaluation of policy proposals such as of non-financial policy impacts; (iii) there is an aggregation problem linked to impossibility to translate from individual preferences to collective choice or from micro assessment results to macro policy recommendations.

In policy impact evaluation public sector's complexity is not taken into account consistently. Revealed difficulties diagnose causes for considerable gap between high theoretical aspirations in policy impact evaluation and its poor practical contribution to consistency of public management. Problems related to three compared evaluation tools are to a large extent comparable and interlinked, which suggests to address evaluation problem in more systematic manner. Impact evaluation difficulties in programming, legislation and budgeting are not independent from each other even though they relate to distinctively different areas and sorts of policy interventions. Programming shapes »our common future«, legislation is framing the uniform norms of social behaviour while budgeting is crucial for every-day life of society. Nevertheless, programming, legislation and budgeting in public sector are functionally dependent areas of public policy intervention. Public budget on the other side in its major part consists of expenditures that are either required by the law or assumed by realisation of programmes. Finally, programming is usually seen in intersection between the need for fulfilment of some legal demands (such as for public utilities, infrastructure) and possibilities for public financing of generally beneficial investments. Finally, legislation defines legal conditions for meeting certain public demands which are decisive in programming and in framing the budget – such as when new environmental standards impose environmental investment from public funds.

Linkages between programming, legislating and budgeting suggest that policy impact evaluation tools that assess their wider social impact should also be somehow operationally linked. All three sorts of policy impact evaluation strategically deal with the same overall system aim of indiscriminate provision of affordable public standard. They use broadly similar evaluation paradigms; they also apply broadly similar set of evaluation criteria and sometimes also share the same assessment data (statistical and evidences produced by government). More systematic evaluation of programs, legislation and budgets would produce more consistent policy advice than separate and independent evaluations. More reliable evaluation results and more consistent policy advice should contribute to improvement in supply of public goods at the same or even lower level of public expenditure.



Three basic evaluation tools, SIA, RIA and PBB can be presented in input-output matrix which show relationships between them and three policy functionalities (programming, legislating and budgeting). For example, sometimes budgeting and programming will trigger indirect impact on society, such as when they induce change in existing normative framework for instance modification of spatial plan which would consequently require “mixed type of evaluation” – regulatory type of program evaluation (impact on opportunities of those affected). Regulatory element of SIA usually covers only formal assessment of concordance of evaluated public program with existing regulation. And the opposite also seems to take place. RIA usually does not study eventual wider impact of new regulation on the society that is otherwise the main general evaluation concern in SIA. In evaluation practice in Slovenia, to which Table 1 refers, this kind of interactions is usually not taking place between SIA, RIA and PBB.

Sign »?« in the Table 1 indicates location of likely systemic gaps in policy impact evaluation, at least as it can be detected from previous analysis presented above. All there evaluation practices seem to cover all three aspects of public policy-making equally incompletely (rows). It seems that the benefit of evaluation is the most fully absorbed for management of the budget (third column). Law-making is part of policy process that is the least supported with evaluations accomplished in other parts of policy process (second column).

This suggests that legal aspect of evaluation in programming and budgeting needs to be strengthened with priority. Contribution of programme evaluation to comprehensive policy advice is found in this cross-sectional view in the middle between RIA and PBB (first column).

Table 1: Relationship between SIA, RIA, PBB and three public sector functions, Slovenia

Effect Cause	Programming	Legislating	Budgeting
Evaluation in Programming (SIA)	SIA (with several difficulties)	? Program impacts on normative system	SIA (financial impacts)
Legislative Evaluation (RIA)	? Impact of legislation on programs	RIA (with several difficulties)	RIA (financial impacts)
Evaluation in Budget preparation (PBB)	PBB (performance monitoring of budgetary items)	? Legal implications of budgeting	PBB (with several difficulties)



---

A working conclusion from provisional schematic presentation in Table 1 and previous discussion is that evaluation can not contribute to public sector's policy integrity also because it is too simplified and in cross-section perspective unsystematic. Paper has identified strong and direct link between policy impact evaluation and in/consistency of policy-making. So it is entirely possible that legislator, program designers and budget manager receive inconsistent policy advice until different evaluation tools are differently »calibrated« in their limited scope and scale of judgement as well as until these tools are tightly related between each other. This suggests that impact evaluation may itself operate as a factor of public sector inconsistency and so a cause of its poorer performance. In this way initial hypothesis is reinforced.

#### 4 Conclusions

Initially stated research hypothesis says that a large deal of problems in public management is not caused by the complexity of public affairs themselves but by the inappropriate address of them with the private sector rationality imprinted in NPM doctrine. Paper has identified close link between NPM doctrine and low policy relevance of evaluation conclusions at least in Continental European conditions. Reductionist approach which is typical for assessment of private sector performance presently prevails in public sector such as when linear cumulative techniques are applied to generate system wide conclusions in impact evaluation. When complexity of public domain is recognised, such simplistic reasoning is not acceptable. Ability for complex assessment of policy impacts depends on developing new synthesis approach. Precondition for this is that evaluator understands social complexity in scale/scope and studies them indiscriminately. This not only justifies the need for the synthesis of detailed evaluation results, but also places aggregation concerns, via neutrality considerations, into the centre of the effort pertaining to better evaluation of public policies.

Evaluators need to investigate more theoretically founded and more interlinked approaches to ex-ante evaluation of proposed public programmes, legislation and budgets. Majority of evaluation studies are accomplished only on one scale and only with two scopes – cause and effect, in this way overlooking complexity of public domain where causes and effects are intertwined. In the future new approaches are needed that are capable of evaluating public programs in their multiple scopes and multiple scales indiscriminately.

Today evaluators try to earn their neutrality in the first evaluation step of neutral impact assessment. Paper showed that neutrality in evaluation must be earned also in the appropriate synthesis of assessment results. As more susceptible to deep differences and more inclusive, the soft synthesis that indiscriminately evaluates

---

policies in multiple scopes and on multiple scales has a greater potential for objectivity (Mertens, 1999).<sup>74</sup> Appropriate summation in evaluation is an effective shield against political influence (Chelimsky, 1995).<sup>75</sup> Paper suggests that neutral synthesis is possible in intersectional or soft rational approach which is developed in mid-range view of weak ties between evaluated social phenomena. Paper also comes to a conclusion that policy impact evaluations should be implemented in different areas as mutually supporting tools, not as a set of unrelated efforts.

Ability to model complexity of social issues in public management appears much more decisive factor for its improved performance than cost-efficiency arithmetic. Even though government is seen as the problem in studying public performance it is not so exclusively in terms of its size and cost but more in terms of its lack of integrative thinking and cohesive performance. More consistent impact evaluation will produce more consistent policy advice and more cohesive policy will save public resources – so to say on the output side of the problem. When policies are so to say »secondary effective«, when they implement projects with strong positive secondary impacts, they can achieve incommensurably different goals without endangering overall system cohesiveness.

Secondary effectiveness is much more important criteria for evaluation of policy impacts than it is currently thought. Various public agencies see some impact of their activity as intended or primary while all other events as unintended or secondary. But when all sectors are taken into account simultaneously, as in strategic impact evaluation, situation inverts. In a society as a whole nothing can be justified as primary for everybody, taking into account social incommensurability of values and knowledge. Something what is treated as primary always belongs to a minority view. On the macro level, for majority of those involved in the evaluation bulk of policy impacts is perceived as impacts of secondary importance which means that secondary impacts prevail and are more important in macro view. When different public policies are equally important and there is no mechanism to install an optimal public policy a policy proposal that has the most favourable secondary impact ought to be chosen (compare with Demsetz, 1969).<sup>76</sup> The yardstick by which public sector's programmes are measured ought to be overlap and secondary policy impacts. The same conception is relevant to the evolutionary social thought of both Hayek and Popper, who take the view that the unintended consequences of action are the principal concern of social science, and, indeed, that the existence of unintended consequences is a precondition for the very possibility of a scientific understanding of complex

---

<sup>74</sup> Mertens D.M. Inclusive Evaluation: Implications of Transformative Theory for Evaluation. Sage, American Journal of Evaluation, 20/1(1999):1-14.

<sup>75</sup> Chelimsky E. Politics, Policy and Research Synthesis. Sage, Evaluation, 1/1(1995):97-104.

<sup>76</sup> In Schnellenbach, 2005.

---

society.<sup>77</sup> In this way secondary concerns are for public domain equally important as primary concerns in the domain of private sector (neo-classical) rationality. As a result, managers in public sector should be much more than presently aware not only of their own agency's or primary sector's objectives narrowly defined, but also of »other« or secondary effects and wider implications that arise from their (in)activity on system wide performance.

---

<sup>77</sup> Vernon R. The »Great Society« and the »Open Society«: Liberalism in Hayek and Popper. Canadian Journal of Political Science/Revue canadienne de science politique, 9/2(1976):261-76, <http://www.jstor.org/stable/3230923>, [III/09]

About the Slovenian evaluation society

Society has been established in 2008 with the vision to affirm independent evaluation of complexity social phenomena. Society operates as a platform of the civil society. Its work is divided between three permanent commissions: for ethical codex and evaluation standards; for meta-evaluation; and for evaluation studies.

About the SES's Working papers series

Working paper series is freely accessible on the internet. It publishes scientific and technical papers from the aspect of evaluation of public policies and from related disciplines. Papers are reviewed and bibliographic catalogued (CIP) ensuring that information is available in all catalogues in Slovenia and elsewhere. Already published in SES/WP:

	<b>Volume 1:</b>
<u>Excercising Aggregation</u> (In Slovenian; B. Radej, 23 pp).	1/2008
<u>Synthesis of Territorial Impact Assessment for Slovene Energy Programme</u> (In Slovenian; B. Radej, 43 pp).	2/2008
<u>Meso-Matrical Synthesis of the Incommensurable</u> (B. Radej, 21 pp).	3/2008
	<b>Volume 2:</b>
<u>Anti-systemic movement in unity and diversity</u> (B. Radej, 12. pp)	1/2009
<u>Meso-matrical Impact Assessment - peer to peer discussion of the working paper 3/2008; report for the period 8/08-2/09</u> (B. Radej, ed., 30. pp)	2/2009
<u>Turistic regionalisation in Slovenia</u> (in Slovenian, J. Kos Grabar, 29 pp).	3/2009
<u>Impact assessment of government proposals</u> (In Slovenian; B. Radej, 18 pp.)	4/2009
<u>Performance Based Budgeting</u> (In Slovenian; B. Radej, 33 pp.)	5/2009
	<b>Volume 3:</b>
<u>Beyond »New Public Management« doctrine in policy impact evaluation</u> (B. Radej, M. Golobič, M. Černič Istenič, 25 pp)	1/2010
<u>Basics of impact evaluation for occassional and random users</u> (In Slovenian; B. Radej, forth.)	(2/2010)

About the authors

**Mojca Golobič** is full professor in spatial planning, environmental planning, landscape conservation and participative planning at graduate and postgraduate programs at the University of Ljubljana, Biotechnical Faculty. Fulbright visiting lecturer at Harvard Graduate School of Design (2003-04) and visiting lecturer at Universities of New Hampshire, Durham, University of Oregon, Eugene and University of Washington, Seattle in USA (2004), and Jelgava University in Latvia.

**Majda Černič Istenič** is associate professor of sociology at the University Ljubljana, Biotechnical Faculty and Senior research fellow at the Sociomedical Institute at the Scientific Research Centre at Slovenian Academy of Sciences & Arts (SRC SASA). Her field of expertise and main research interests are rural sociology and population dynamic. She participated in many domestic research projects as a leader or as a member and also in several European Commission projects in the 5<sup>th</sup>, 6<sup>th</sup> and 7<sup>th</sup> Framework Programme. She was also the evaluator of European Commission Marie Curie Programmes (2007-2009).  
Member of Slovenian Evaluation Society.

**Bojan Radej** is a methodologist in social research from Ljubljana. Master degree in macroeconomics, University of Ljubljana - Faculty of Economics (1993). Professional Experience Record: Governmental Institute of Macroeconomic Analysis and Development (1987-04; under-secretary to the government), areas of work: sustainable development (1998-04), chief manager of the modelling department (1993-5); initiator and the first editor of Slovenian Economic Mirror (1995-8); Co-Editor of IB journal (2001-04).  
Member of Slovenian Evaluation Society.

Published by Slovenian Evaluation Society ([Creative Commons](#)).

Contact address - editor: [info@sdeval.si](mailto:info@sdeval.si)

Design of the SDE logo: Naja Marot, UIRS, [naja.marot@uirs.si](mailto:naja.marot@uirs.si)

SDE is supported by Inštitut za ekonomska raziskovanja, Ljubljana, <http://www.ier.si/>