



THE IMPACTS OF THE NEW PUBLIC  
MANAGEMENT IN EUROPE:  
A META-ANALYSIS

**COCOPS WORK PACKAGE 1 – DELIVERABLE 1.1**

Christopher Pollitt

Sorin Dan

14 December 2011



Funded under Socio-economic Sciences & Humanities

© COCOPS 2011

### **About COCOPS**

The COCOPS project (Coordinating for Cohesion in the Public Sector of the Future) seeks to comparatively and quantitatively assess the impact of New Public Management-style reforms in European countries, drawing on a team of European public administration scholars from 11 universities in 10 countries. It will analyse the impact of reforms in public management and public services that address citizens' service needs and social cohesion in Europe. It is funded under the European Commission's 7th Framework Programme as a Small or Medium-Scale Focused Research Project (2011-2014).

### **About the Authors**

Christopher Pollitt is BOF/ZAP Research Professor of Public Management at the Public Management Institute, Katholieke Universiteit Leuven.

Sorin Dan is Researcher at the Public Management Institute, Katholieke Universiteit Leuven.

We would both like to acknowledge the invaluable input to this study made by the IT officer at the Public Management Institute, Anita van Gils.

The research leading to these results has received funding from the European Community's Seventh Framework Programme under grant agreement No. 266887 (Project COCOPS), Socio-economic Sciences & Humanities

ISSN 2211-2006

## THE IMPACTS OF THE NEW PUBLIC MANAGEMENT IN EUROPE: A META-ANALYSIS

### **Abstract**

In order to assess the impacts of the NPM we first define NPM and then set out what we mean by impacts. Subsequently we report our analysis of our database of nearly 520 relevant studies (the database can be accessed through the COCOPS website at [www.cocops.eu](http://www.cocops.eu)). Our main findings are, first, that only a minority of studies deal with outputs, and a very small percentage with outcomes. Most studies of NPM confine themselves to ‘internal’ changes in the activities and processes of public organizations. Subsequently we focus on the minority of studies that do report on outputs and outcomes. Whilst the largest proportion of these do register performance improvements, very substantial percentages report no change or even an actual worsening of performance. This high degree of variation appears to be connected with a number of contextual influences, so in the later sections of the paper we analyse the different types of contextual factors that seem to be important. Finally, we offer informed speculation on why NPM reforms have been so widespread when their evidence base appears so fragile.

**Keywords:** New Public Management (NPM), evaluation, outcomes, outputs, context, Europe

## 1. Introduction

This meta-analysis is the first deliverable of the first Work Package of the EU Seventh Framework programme project *Co-ordinating for Cohesion in the Public Sector of the Future* (hereafter COCOPS – see [www.cocops.eu](http://www.cocops.eu)). The task we were set was to produce a *meta-analysis of the impacts of New Public Management (NPM) reforms across the EU*. In this paper, therefore, we will first set out what we mean by (respectively) NPM, impacts and meta-analysis, and then go on to review the substance of our findings.

The database we have constructed is, we believe, novel in at least two respects. First, it is relatively large – almost 520 studies. We do not know of any other databases where such a substantial number of NPM studies have been both listed and systematically analysed in some detail. Second, our concentration on *Europe* is unusual: until now the bulk of Anglophone literature on NPM has covered Anglophone countries, especially Australasia, the UK and the USA. Now we have a much clearer picture of what has been happening across continental Europe (as well as the UK and Ireland).

The definitional exercise on which the database is founded was anything but simple. The interpretation of our results rests significantly on a number of significant definitional and operational choices that we made. (Any serious researcher on this subject would be obliged to make such choices, in one way or another.) These preliminary categorizations are discussed in sections 3-7 below. The later sections of the report (8-13) set out what, using these definitions and categories, we found. We distinguished between impacts on *processes (or activities)*, *outputs* and *outcomes*. Within those broad categories we have paid particular attention to impacts on efficiency, effectiveness, and the attitudes of those who use public services. We have also looked for evidence of impacts on *social cohesion*, which for our particular purposes we have defined as having to do with equity of access to services and with the solidarity and commitment of public servants themselves. In the final section (13) we offer some concluding thoughts and informed speculations.

Our main method has been to collect and analyse a very large number of studies – both academic and practitioner – which appear to address the issue of the impacts of NPM. Initially, we tried not to discriminate by method or approach (so that we can hear a variety of ‘voices’ and maximize the size of our total stock of material). Later in the analysis, however, we begin to make some comments on issues of methods and strength of evidence.

Our 520 studies – from right across Europe – have become part of a database we have created and made available through the COCOPS website at [www.cocops.eu](http://www.cocops.eu). The database can be sorted according to the type of document, country, sector and year. The studies themselves can be downloaded through the database in cases when they are freely available. In the case of studies that are not freely available, such as some books and journal articles, we have provided a link to the respective websites where these sources can be purchased or where an abstract or summary can be consulted.

We have used a Microsoft® InfoPath form created by the COCOPS Work Package 1 team, built into Microsoft® SharePoint. This ITC tool enabled the COCOPS team members at 11 universities in 10 different countries to supply the studies, and to code and share them within the project. In a second phase the SharePoint system facilitated the analysis of the data by generating a series of descriptive statistics (included in this paper, section 9). The built-in InfoPath form was created with the option of indicating whether a particular observed impact meant that the specified ‘targets’ improved, remained unchanged or deteriorated. This therefore helped to generate a count of the number of entries indicating that processes or activities, outputs and outcomes improved, worsened or remained unchanged (section 10).

## **2. Defining NPM**

Defining NPM is the first step but in itself is not at all easy. In its origins it is strongly associated with UK Prime Minister Thatcher and US President Ronald Reagan, and with the New Zealand Labour government of 1984. Neither Mrs Thatcher nor Ronald Reagan were any friends of the ‘planning’ approach which had been the orthodoxy in the US and UK public sectors of the 1960s and early ‘70s. During their periods in power in the 1980s they, and many of their advisers, favoured what they considered to be a more ‘business-like’ approach. Gradually, partly through doctrine and partly through trial and error, this general attitude crystallized into a more specific set of recipes for public sector reform. By the early 1990s a number of influential commentators appeared to believe that there was one clear direction – at least in the Anglophone world. This general direction was soon labelled as the New Public Management (NPM) or (in the US) Re-inventing Government (a seminal article here was Hood, 1991). A pair of American management consultants, who wrote a best-seller entitled *Reinventing government* and then became advisers to the US Vice President on a major reform programme, was convinced that the changes they saw were part of a global trend. They claimed that ‘entrepreneurial government’ (as they called it) was both worldwide and ‘inevitable’ (Osborne and Gaebler, 1992, pp. 325-328). At about the same time the Financial Secretary of the UK Treasury (a junior minister) made a speech claiming that the UK was in the forefront of a global movement:

“All around the world governments are recognising the opportunity to improve the quality and effectiveness of the public sector. Privatisation, market testing and private finance are being used in almost every developing country. It’s not difficult to see why.” (Dorrell, 1993)

The increasingly influential Public Management Committee of the OECD came out with a series of publications that seemed to suggest that most of the developed world, at least, was travelling along roughly the same road. This direction involved developing performance management, introducing more competition to the public sector, offering quality and choice to citizens, and strengthening the strategic, as opposed to the operational role of the centre (see e.g. OECD, 1995). Whilst it is now fairly clear that the whole of the world was *not* following the same path (Pollitt and Bouckaert, 2011) it remains true that NPM ideas spread very widely, and are often still seen as the most obvious route to modernization.

There have been many definitional disputes and ambiguities about exactly what the key elements of this widespread trend were supposed to be: 'There is now a substantial branch industry in defining how NPM should be conceptualised and how NPM has changed' (Dunleavy et al., 2006b, p. 96; see also Barzelay, 2000; Gow and Dufour, 2000; Hood, 1990; Hood and Peters, 2004). For the purposes of this meta-analysis we will assume (like Dunleavy) that the NPM is a two-level phenomenon. At the higher level it is a general theory or doctrine that the public sector can be improved by the importation of business concepts, techniques and values. This was very clearly seen, for example, when the then US Vice President personally endorsed a popular booklet entitled *Businesslike government: lessons learned from America’s best companies* (Gore, 1997). Then, at the more mundane level, NPM is a bundle of specific concepts and practices, including:

- Greater emphasis on ‘performance’, especially through the measurement of outputs
- A preference for lean, flat, small, specialized (disaggregated) organizational forms over large, multi-functional forms (e.g., semi-autonomous executive agencies – see Pollitt et al., 2004)
- A widespread substitution of contracts for hierarchical relations as the principal coordinating device
- A widespread injection of market-type mechanisms (MTMs) including competitive tendering, public sector league tables, performance-related pay and various user-choice mechanisms

- An emphasis on treating service users as ‘customers’ and on the application of generic quality improvement techniques such as TQM (see Pollitt, 2003: chapter 2)

Dunleavy et al. have usefully summarized this as ‘disaggregation + competition + incentivization’ (Dunleavy et al., 2006a). Roberts (2010) has added a further interpretive twist by characterizing many NPM reforms as exhibiting ‘the logic of discipline’. By this he means a way of thinking about the organization of government functions that has two major components:

- Scepticism about conventional democratic politics because it tends to produce short sighted, unstable and self-interested policies
- Optimism that if certain subject or activities are legally removed from everyday politics (i.e. if elected politicians are ‘disciplined’ so that they cannot constantly intervene and interfere) then more stable and farsighted policies will be possible

Thus, for example, central banks should be given independence. Executive tasks should be hived off to contract-like agencies, local economic development should be removed from the responsibility of elected local councils and given to special-purpose businesslike bodies. Roberts says that this liberalization agenda revolved around ideas of depoliticization, autonomization and discipline (2010, p.140). Experts and managers will (it is assumed) generate more stable, prudent and realistic policies and decisions than short-term, electorally hypersensitive politicians.

It would be quite mistaken to assume that this formula was internally consistent. As a number of commentators have noted, there is some tension between the different intellectual streams that feed into NPM, particularly between the economistic, principal-and-agent way of thinking, which is essentially low trust, and the more managerial way of thinking which is more concerned with leadership and innovation – and more trusting of the inherent creativity of staff, if only they are properly led and motivated (Boston et al., 1996; Christensen and Lægreid, 2001; Pollitt, 2003, pp. 31-32). The former stream emphasized the construction of rational systems of incentives and penalties to ‘make the managers manage’. The latter emphasized the need to ‘let the managers manage’ by facilitating creative leadership, entrepreneurship and cultural change. Another (related) problem in the formula was the idea that management would have more freedom and yet, somehow, politicians would simultaneously achieve more control (usually this happy result was envisaged from the installation of a more transparent system of performance targets and indicators that would show the political leaders how well things were going, so that they could, if necessary, take

corrective action). Other writers drew a distinction between ‘hard’ and ‘soft’ versions of NPM (Ferlie and Geraghty, 2005). The hard version emphasized control through measurement, rewards and punishment, while the soft prioritized customer-orientation and quality, although nevertheless incorporating a shift of control away from service professionals and towards managers. This seems to map quite closely onto the low trust/high trust tensions mentioned above. As for the ‘logic of discipline’ Roberts comes to the conclusion that, in many cases, depoliticization turned out to be anything but politically neutral, organizational autonomy proved elusive (rather organizations swapped one framework of control for another) and ‘discipline’ was interpreted differently in different cultures (Roberts, 2010, pp. 140-146).

In most of the pro-NPM literature it was assumed that the application of ‘business methods’ would result in a public sector that was cheaper, more efficient, and more responsive to its ‘customers’. These were the key aspirations of the NPM movement, and we will see later in this report how far the evidence of impacts suggests that such hopes were realized.

### **3. Defining the domains of NPM reforms**

For practical reasons of research capacity the COCOPS team decided at an early stage not to try to pursue every NPM reform in every EU country at every level. We limited our search in various ways:

- a) We excluded acts of outright privatization from the Work Package 1 meta- analysis (although privatization will receive some attention in some of the later work packages). We define privatization as the sale of publically-owned assets to the private sector. Therefore corporatization within the public sector remains within our scope, as does contracting out and the creation of autonomous state agencies but not the sale into private ownership of public utilities such as electricity or water supply
- b) We decided to concentrate on central government reforms. We have not attempted to cover reforms of sub-national tiers of government, although here and there, selectively, we included studies of local government if they offered a particularly clear account of the impacts of NPM reforms. In practice, however, the distinction between central and local is less than clear-cut. Quite often reforms will be designed or mandated by central government but then imposed on local government to carry out. Furthermore, in many countries, central government actually controls services which are delivered locally (such as the National Health Service in the UK or schools in France). We have dealt with this ambiguity in a fairly pragmatic way. We have included studies of central government-initiated reforms which have had their prime effects at local levels

- c) We dealt with social cohesion in a very narrow way. Basically, we focused only on two aspects. First we noted any evidence of how reforms might have influenced equity (especially equity of citizen access) in public services. Second, we also paid attention to any elements in the documentation which reported on changes in public service staff morale or attitudes towards their work. We see such changes as at least connected to the idea of cohesion, in the sense that, if morale plunges downwards it is reasonable to assume that the social cohesion of the public service is adversely affected. Equally if, by contrast, reforms cause a marked increase in positive attitudes among staff, then that should be a factor in favour of organizational cohesion within the public service. We should note, however, that positive attitudes among staff are not necessarily or invariably correlated with organizational cohesion. So this is a very tentative exploration. Social cohesion is a complex and slippery concept (Chan et al., 2006) and will be addressed in a more extensive way by some of the later work packages of the COCOPS project
- d) We treated some states more intensively than others. Table 1 (below) shows the level of scrutiny we gave to the different states. This was determined by a variety of factors, including the knowledge, linguistic abilities, and personal networks of the partners in the COCOPS team. Obviously it was easier to survey the literature of a country where a COCOPS partner was based, and obviously it was much more difficult in countries where none of our partners had command of the language (e.g. Greece). Our coverage, however, was equally shaped by some theory-related principles of selection. We aimed at – and achieved – coverage in depth of at least two countries in each ‘cultural group’. By cultural group we mean countries that shared important political histories and norms – elements that continue to shape their approach to public administration. The main groups we identified are all widely recognized as such in the academic literature, viz. the Nordic group, the ex-Napoleonic Mediterranean states, the CEE and the ‘big three’ (France, Germany and the UK – each of which has developed a strong style of its own, which has influenced other countries closely connected with it). Further discussion of these groupings can be found in e.g., Lynn, 2006; Ongaro, 2009; Painter and Peters, 2010.

**Table 1. Level of coverage of NPM reforms in COCOPS Work Package 1**

<b>State</b>	<b>Level of treatment in WP1</b>
Austria (AT)	In depth
Belgium (BE)	In depth
Bulgaria (BG)	Not covered
Croatia (HR)	Medium

---

Cyprus (CY)	Not covered
Czech Republic (CZ)	Medium
Denmark (DK)	Medium
Estonia (ES)	In depth
Finland (FI)	In depth
France (FR)	In depth
Germany (DE)	In depth
Greece (GR)	Light
Hungary (HU)	In depth
Ireland (IE)	Light
Italy (IT)	In depth
Latvia (LV)	Light
Lithuania (LT)	Medium
Luxembourg (LU)	Light
Malta (MT)	Not covered
Netherlands (NL)	In depth
Norway (NO)	In depth
Poland (PL)	Light
Portugal (PT)	Medium
Romania (RO)	In depth
Slovakia (SK)	Medium
Slovenia (SI)	Not covered
Spain (ES)	In depth
Sweden (SE)	Medium
Switzerland (CH)	Medium
United Kingdom (UK)	In depth
European Commission (EC)	Medium

---

We should also note that defining what is and what is not a public organization is by no means always straightforward (see for example, the complexities encountered in CEEP, 2010 or Clifton and Díaz-Fuentes, 2010: section 2). By focusing principally on central government bodies we hope we have, however, avoided the worst of this.

Finally, it is important explicitly to state that NPM reform and public management reform are by no means the same thing. It is necessary to make this pronouncement because there has occasionally been a tendency, perhaps especially among continental European commentators, to see anything that is labelled ‘management reform’ in the public sector as being some species or aspect of NPM. We find this loose and confusing – to paraphrase Aaron Wildavsky, if NPM is everything, then it is nothing. Our position is quite clear: a) NPM

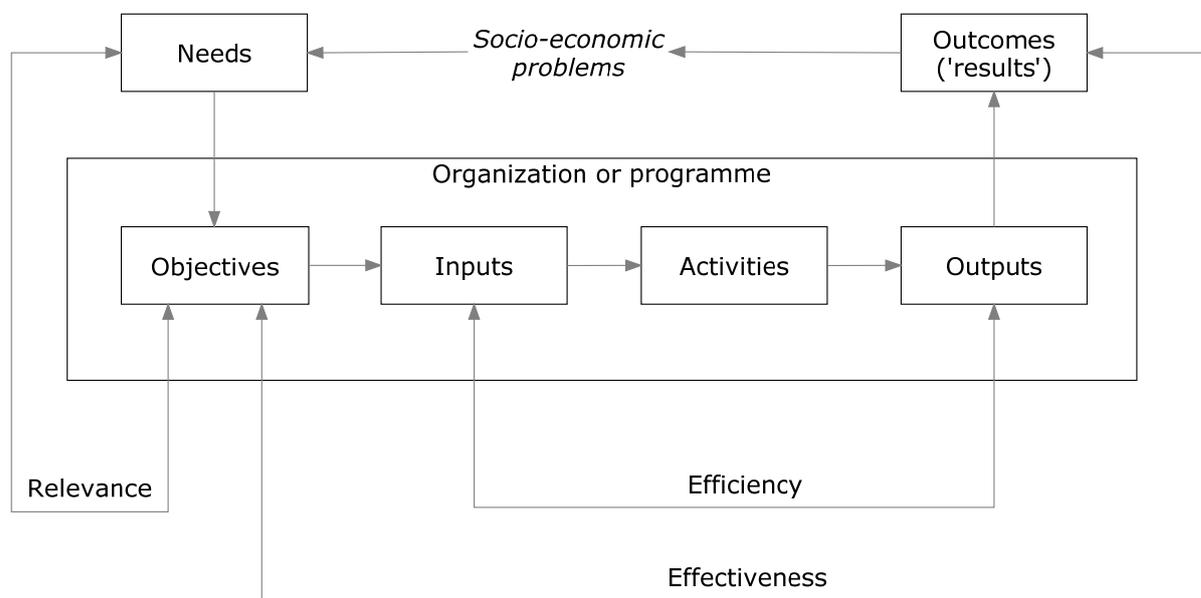
reforms partake of several or all of the characteristics identified above, b) there are many other types of reform going on which cannot be classified as NPM (for example, ‘joined-up government’ or collaborative networking) and c) very important though these other types of reform undoubtedly are, this paper is not concerned with them. Our focus is on the impacts of NPM.

#### 4. Assessing impacts

The COCOPS research proposal committed Work Package 1 to a ‘meta-analysis of the impact of NPM on efficiency, effectiveness, quality and social cohesion’. Almost every term in this short phrase has led us into extensive internal debate, not least the notion of ‘impacts’.

‘Impacts’ are things that can reasonably be supposed to have been *caused* by the reforms. They could be things very close to the reforms (for example, staff lost their jobs because some activities were contracted out) or much ‘further out’ (young people got better jobs because the quality of their education went up, because the performance of the schools they attended was closely measured and publicized, so that teachers were obliged to improve the standard of instruction – quite a long causal chain). In thinking about impacts we adopted a fairly standard model of policymaking (see Figure 1, below).

**Figure 1. Performance: A conceptual framework**



(Adapted from Pollitt and Bouckaert, 2004, p. 13)

Although the conceptualization and the terms in Figure 1 are, as we just said, fairly standard, and are used by many academics, some national audit offices and some government studies all alike, our repeated experience is that they are nevertheless very variably interpreted

elsewhere. In some cases normatively positive terms like ‘efficient’ or ‘effective’ are stretched in order to make an argument sound better. For example, government reports may claim an increase in ‘effectiveness’ when what they actually mean is that organizational staff had adopted better working methods – which in terms of Figure 1 would be just process changes, and nothing to do with outcomes (for an example of this see PA Consultants, 2002). Politicians, in particular, splash words like ‘efficient’ and ‘effective’ around like fresh paint, but some academics are also fairly loose in the way they use these terms.

We therefore need to say a few words about the relationships conceptualized and represented in Figure 1, and how we intend to use these terms. First, an organization or programme is conceived as a set of *activities or processes*. These would include organizational arrangements like the division of responsibility, the allocation of authority, the standard operating procedures, and so on. These procedures enable the organization to deploy and redeploy its resources (staff, money, buildings etc) which are collectively termed *inputs*. From these activities and processes the organization or programme then produces a set of *outputs*, which could, for example be lessons (in a school), licenses (from a licensing agency), medical treatments (from a hospital) and so on. These outputs are, in a sense, what the organization ‘gives’ to the outside world – to citizens, to civil society associations and to business firms. They are like messages, passing across the membrane that separates (on the one hand) the state from (on the other) the market sector and civil society. Outputs are invariably intended to produce desirable *outcomes* beyond the organization or programme – so school lessons are supposed to produce educated students and hospital-provided medical treatments are supposed to produce the cure or alleviation of ill-health (see also COCOPS Work Package 6, ‘NPM and social cohesion’). *An outcome is something that happens in the world outside the organization and the programme: it is an effect ‘out there in the real world’.* Some analysts make distinctions between ‘intermediate outcomes’ and ‘final outcomes’, but we decided that such a distinction, though useful in some contexts, would be of limited value in our field of study, where outcomes measures of any kind are rather rare.

One final point to add about effectiveness is that one measure that is potentially very useful to policymakers is that of *cost effectiveness*. It is not shown in figure 1 because it is essentially the same relationship as shown for basic effectiveness, i.e. the degree of match between policy goals and outcomes. However, cost effectiveness adds a unit cost calculation, i.e. how much of the goal is gained for each unit of expenditure. Thus, for example, a measure that showed how many cancer deaths were avoided per £1m spent on a cancer screening programme would be a cost effectiveness measure (‘an additional £10M would save an

additional 14 lives’, for example). Unfortunately, there are very few cost effectiveness measures in the field of public management reform.

The performance of organizations and programmes (the value of their activities) is usually thought of in terms of certain relationships between these inputs, outputs and outcomes. Thus the ratio between inputs and outputs is a measure of efficiency (or ‘technical’ efficiency or ‘X-efficiency’, or ‘productivity’). If you can get more outputs for the same inputs, you have achieved an efficiency gain. If you can maintain steady outputs while reducing inputs you have also achieved an efficiency gain – in both cases the ratio between inputs and outputs improves. Effectiveness, however, is a different concept, which is usually conceived as the degree to which the outcomes match the original goals or objectives set for the organization of programme. So, if a hospital is created with the goal of reducing premature deaths from heart disease, and it performs operations and delivers treatments which do indeed reduce the incidence of premature deaths within its catchment population, then it has been effective.

We should also mention *economy*, which we define as the simple reduction or lowering of inputs – in a word, cost reduction. We need to keep this in the frame both because pro-NPM reformers often claimed that their changes would reduce costs and because a number of the studies in the database comment specifically on this aspect. Economising may not be quite as ‘catchy’ as increasing efficiency or improving effectiveness but, particularly in our current climate of fiscal austerity, it is often a prominent and unavoidable ‘fact of life’ for politicians and public servants alike.

As many writers have remarked, if goals are multiple, conflicting or ambiguous then it will be difficult to determine effectiveness which will, in effect, become a ‘contested concept’. Unfortunately for the analysts, policy goals frequently *are* multiple, conflicting or ambiguous, not least because that is what politicians may need to get sufficient agreement to launch the policy in the first place. This certainly includes public management reforms which are often claimed to be all things to all men – to save money, raise service quality, increase effectiveness, etc, etc. The studies we have embraced in the database do indeed comment on many kinds of goals and effects – for example on power relations, political steering, accountability, legitimacy, transparency and other issues. We cannot claim to have given each and every one of these equal attention. Our main focus has been on what we identified earlier as the key aspirational goals of NPM – a cheaper, more efficient government that was also more responsive to its ‘customers’. However, we do mention some of the other goals and effects where they seem particularly significant (see also section 10).

Although Figure 1 depicts a helpful and widely-used conceptual framework it also raises a number of questions. One is – where does public opinion fit into all this? Our answer is that we *have chosen to treat changes in public opinion (e.g. changing levels of satisfaction or trust) in respect of public services as an outcome*. Clearly it is not an output, because it is not ‘made’ by the organization or programme – rather it is a reaction to or assessment of that output (among other things). However, it is a different kind of outcome from some other outcomes. Whereas there is something concrete and clearly countable about, say, reductions in the number of road deaths or increases in the percentages of students passing a national test, satisfaction and trust appear to be more subjective – and volatile. That does not make opinions any less important but it does imply that analysts will need to treat such data with considerable interpretive caution. Note that citizen reactions to public services, although *in themselves* a kind of outcome, may refer to any stage in the policy process. In other words, the public may have strong feelings about cuts in health service budgets (which are inputs) or about improved queuing systems in the benefits payment office (which is a process change) or about new kinds of school lessons (which are outputs) or about the way in which the environmental health service has allowed a local river to become visibly polluted (which is an outcome).

Note also that effectiveness and efficiency by no means always move together. It is perfectly possible to carry through reforms which improve effectiveness but which do not affect efficiency, or which even reduce efficiency. Similarly, it is wholly conceivable that one can make changes that will increase efficiency, but also lead to some loss of effectiveness. For example, if a university hires top professors instead of junior lecturers to teach its courses, the effectiveness of its teaching may go up, but, *ceteris paribus*, with a higher salary bill and only the same number of lessons, its efficiency will go down.

Another point to be borne in mind is that, increasingly, official reports and evaluations, as well as academic studies, make use of complex indices of performance, which combine two or more elements (see e.g. Arndt, 2008; Audit Commission, 2009; Pollitt, 2011; Van de Walle et al., 2008). These aggregated indices can be very useful to busy decision makers or to non-specialists and citizens, but they may also (deliberately or inadvertently) conceal underlying trades-off between two or more desirable values (e.g. equity versus efficiency). They can also give spurious precision to judgments which are more correctly seen as hedged about by quite wide brackets of uncertainty (Jacobs and Goddard, 2007). Furthermore, from our point of view they pose classification problems. In Audit Commission 2009, for example, local authority performance is classified as ‘excellent’, ‘good’, ‘poor’ (etc.) where each of these categories is in practice a complex aggregation of measures of output, assessments of

strategic planning and capacity in the authorities concerned, and a number of other elements. How then do we classify a shift from, say, ‘good’ to ‘excellent’? It could be due to an increase of effectiveness, or efficiency or strategic planning capacity, or any combination of these things. We have no short solution to this, and have merely attempted to pick out from such aggregates the separate elements – where we can.

The above approach to impacts can operate on very different levels. One study may deal with the impacts of a particular organization, another with the impacts of a reform programme which embraces several or many organizations, and a few studies actually try to capture changes right across the civil service or even the entire public sector. When we come to analysis of the contents of the database we will usually need to distinguish between these different kinds of study. In the section on contextual influences (10) for example, we distinguish between influences that operate on an international scale, a national scale, within a given sector, or just on a particular organization or department.

Apart from process changes, efficiency and effectiveness, our terms of reference required us to pay attention to two dimensions of performance not mentioned on figure 1 – ‘quality’ and ‘social cohesion’. Both terms appear quite frequently in policy as well as academic debates, but the sad truth is that both terms are even more subject to multiple and vague definitions than ‘efficiency’ and ‘effectiveness’. ‘Quality’ is a notorious term, in that its positive normative connotations mean that it has been used to cover almost any aspect of public services. It may be used to cover the views of service producers and experts, or the opinions of service users/consumers. The well-known reform approach known as Total Quality Management (TQM) rests heavily on the opinions of consumers to define quality (although some of these consumers may be ‘internal customers’ rather than the citizen in the street). ‘Quality’ has certainly been used to refer to all manner of combinations of processes, outputs and outcomes – which is one reason why it can be a source of confusion (Pollitt and Bouckaert, 1995: chapter 1). We therefore treat it as a multi-dimensional concept, the actual dimensions of which need to be defined in each particular context. Thus (for example) a study showing that a Total Quality Management (TQM) system had been installed at government organization X would count as a process change. (Installation alone does not guarantee a change in outputs, still less outcomes.) A deeper study, showing that service users at X had noticed faster more accurate service since the TQM system had been installed would count as an output. And a further study showing that the faster and more accurate service from X had led to healthier patients or catching more criminals (or whatever) would count as an outcome.

As for social cohesion, we have here another slippery, multi-meaning concept and one which, like ‘quality’ carries a distinct normative charge (Chan et al., 2006; Green and Janmaat, 2011). ‘Cohesion’, like ‘quality’, presents as A Good Thing. It is a term much used within EU policymaking, but not nearly as prevalent in the Anglo-Saxon world. It usually has to do with avoiding or ameliorating gross socio-economic inequalities, on the grounds that such imbalances are both ethically and socially corrosive. Such inequalities can occur in a vast variety of forms (for only some of these, see Kearns and Forrest, 2001). They may be geographical (some regions or localities are left behind in development) or economic (an increase in income or wealth inequalities) or social (some parents are able to buy a high quality education for their children while others have no choice but to send theirs to substandard slum schools). So the meaning of social cohesion tends to shift from one context to another (Green and Janmaat, 2011). Given our focus on NPM reforms, we decided to concentrate primarily on issues of access to public services. Did the reforms we are interested in generate increased inequalities in citizen access, or reduced inequality, or did they leave things more or less the same? It has to be said, however, that whilst there are a few important studies which examine this aspect (e.g. Boyne et al., 2003), most NPM literature is far more concerned with efficiency than with equity. Nevertheless, we look at what there is. A second aspect of cohesion that we pay some attention to is that of the attitudes and morale of public service staff. As we said above, we see falling morale as a possible indicator of some loss of cohesion within public sector organizations, and rising morale as a sign that cohesion may be increasing. Some of the reports we study do contain evidence of changes in morale (or related concepts) so we will include that in our analyses, although we are conscious that these connections (morale-cohesion) are by no means certain or automatic. Finally, with respect to figure 1, it should be noted that (again, like quality) threats to social cohesion can occur at any stage in the policy process – inputs, processes, outputs and/or outcomes. The *effects* of lowering cohesion – the ethical problems and the social corrosion – will, however, be regarded, whether intended or unintended, as a particular class of outcomes.

Figure 1 does not mention ‘impacts’ as such, so there is still a need to choose how widely to construe the notion. One could take a very tight, puritanical line and say ‘only outcomes should be regarded as impacts’. Or one could be more liberal and also allow outputs. Or one could adopt an even broader definition and include reform-induced changes in activities (processes) as well. To stretch even further back up the chain, and include inputs, seemed to us a step too far (if one puts more money into contracting out and less into in-house services, is that an ‘impact’ of the contracting out reform?). In the end we adopted the following guideline: *we will consider changes in outcomes, outputs and activities/processes as impacts, but will devote more analytic attention to outputs and outcomes than to activities/processes.*

The way we collected the data allowed us to distinguish studies in each of these three categories (i.e. studies which focus on activities/processes, output studies and outcome studies). Some studies obviously include data on more than one of these categories.

This is not by any means the end of the analytic problem. There is also the vexed issue of *attribution*. To be the impact of an NPM reform the outcome, output, or change in process, must appear to be the *result* of that reform, not of other developments which may be happening at the same time. There must be a plausible causality. This condition is, however, often hard to satisfy (in fact in the literature we have examined it frequently is not satisfied at all). A classic example would be shifts in public satisfaction, or trust, in government or in the civil service. If there is a shift (either upwards or downwards) it *could* be the result of a preceding public service reform. However, it *could* also be the result of a general well-being factor, linked to an economic boom, or to ephemeral shifts in party politics (a new leader, a scandal) or to pre-existing personal expectations or hopes, or any combination of these. Tracing shifts in public opinion to specific reforms can be very hard to do (Clifton and Díaz-Fuentes, 2010; James, 2010). Shifts in trust may, in fact, be the result of reforms in some instances but an important causal factor in others. The problems of attribution are therefore widespread, but tend to become particularly acute when *outcomes* are being considered, mainly because there are so many other changes going on out there ‘in the world’ that to be sure that the reform was the exclusive, or at least the predominant cause of something is frequently extraordinarily difficult.

The issue of causality is, of course, both fundamental and controversial. Some species of contemporary social sciences will hardly admit the notion of causality at all. Others insist that causality is central to social science explanation, but adopt quite a restricted notion of what will count as causality (Pollitt and Bouckaert, 2009: chapter 9). In this work we broadly follow Kurki (2008, pp. 210-234) in accepting a multi-dimensional concept of causality, in which material causes, formal causes, efficient causes and final causes can all play interacting roles in an explanation.

All in all, therefore, it can readily be appreciated that assessing the impacts of NPM reforms is very far from straightforward. It involves several layers of categorization and interpretation, each of which has some effect on what is eventually ‘found’. Given our terms of reference, this is unavoidable. Terms like ‘quality’, ‘social cohesion’, and even ‘efficiency’ and ‘effectiveness’ are not stable, technical entities. They have entered common – or at least bureaucratic and political – speech, and have there been deployed in a fantastic variety of ways. Since COCOPS is very much engaged with the practitioner as well as the academic

literature, we are obliged to deal with all this material, however messy or inconsistent it may sometimes be.

More generally, it is important to understand that the body of work contained within the database derives from a mixed, increasingly multinational community, consisting of academics, public servants, management consultants and politicians. It is therefore unsurprising that the reasons for becoming engaged with NPM reforms have differed. Some participants want to find the best way forward – reforms that will work to solve some real (or imagined) problems. Some want to justify a recent choice of a new direction – to defend a new policy against attacks from the political opposition or criticism from the media. Some wish to package and sell sets of ideas (‘best practice’, ‘the reinvention model’, etc.). Management consultants, ‘experts’ and governments all do this (e.g. Federal Ministry of the Interior, 2006; Kaufmann et al., 2007; Osborne and Gaebler, 1992; Prime Minister and Minister for the Cabinet Office, 1999). Some hope to sound progressive and look good at an election. And some – mainly the academics – simply want accurately to describe and explain what is happening or has happened in the world of reform.

This mixture of motives means that the dividing line between descriptive and analytical (‘is’) statements and normative (‘ought’) statements is frequently hard to find. The desire to understand and explain is often tangled up with the desire to promote and support a particular kind of reform. Those reading the literature therefore need to be especially sensitive to the likely interests of the author(s), to unspoken assumptions, to the strength of evidence in relation to the size of the claims being made, and so on. This is what used to be called ‘source criticism’ and it is a vital technique for those who wish to investigate the literature on NPM. For example, a someone who researched NPM solely by visiting government websites would be likely to come up with a picture of what was going on that was both over-simple and over-optimistic. Even texts produced by academics cannot be assumed to be ‘neutral’, partly because many public management academics also work in consultancy and advice roles but also because the academic world is itself divided between competing theoretical and methodological camps (Pollitt and Bouckaert, 2009: chapter 9).

## **5. Actual selection of studies for uploading to the database**

Studies were identified both by our own literature review and by calling on the expertise of our COCOPS partner institutions, and of a number of other collaborators who kindly volunteered to search out material from their own countries.

As for our own literature review, we began by checking everything that had been published since 1980 in what we considered to be the seven of the leading mainstream Anglophone public administration journals (they are all listed in the Thompson Index/SSCI), plus a few others:

- Journal of Public Administration Research and Theory (JPART)
- Public Administration
- Governance
- Public Administration Review
- International Review of Administrative Sciences
- International Public Management Journal
- Public Management Review
- Public Policy and Administration
- Evaluation

In addition to these academic sources we trawled a number of promising organizational publication lists, including those of:

- OECD
- SIGMA
- World Bank
- National government websites on central government reforms (in those countries where they are available)
- National audit office performance audits (in those countries where they are available)

Alongside this literature search, as mentioned above, we worked either directly or through our COCOPS partners with a number of collaborators. From these sources we gained citations of official publications in the countries concerned, as well as academic publications in non-Anglophone journals. The COCOPS partners were in Estonia, France, Germany, Hungary, Italy, the Netherlands, Norway, Spain and the UK. The volunteer collaborators were mostly in CEE and the Baltic States but also in Italy, Portugal and Switzerland. This opportunity provided us with wider access to studies and better insight into the specifics of the evaluation of impacts of NPM reforms in the respective states, but it also challenged our attempts to standardize the coding and analysis process. It also showed us even more clearly how slippery the concepts that we operated with were and how challenging was the goal of assessing the impacts of NPM reforms.

We developed a set of guidelines and definitions of key terms for our collaborators, who then provided the documents and a summary of key findings as they pertained to impacts.

In a second stage, either we or our COCOPS partners checked the documents and data for accuracy and relevance and uploaded them into the SharePoint database those that met the criteria for impact assessments of NPM-type reforms (see introduction for a more detailed description of the database). In rare cases, some of the collaborators uploaded the documents directly into the database. The slipperiness of the interpretive problems necessarily involved in studying NPM impacts meant that, despite extensive discussions of the foregoing issues across all the partners in the COCOPS research team, and the development of explicit guidelines for study selection, when it came to actually uploading selected documents into the database, some inconsistencies appeared. For example, one partner uploaded a paper entitled ‘E-government and the transformation of public administrations in EU countries: beyond NPM or just a second wave of reforms?’ (Torres et al., 2005). The title certainly sounded relevant, but closer inspection revealed that a) the paper dealt solely with sub-national governments and b) that its relation to NPM as we have defined it was vestigial (it took a somewhat ambiguous position, arguing that both that e-government was a kind of successor to NPM, but also that it was a potentially transformatory route to something regarded as quite different, namely ‘governance’). Another example would be studies that discussed the logical coherence or underlying philosophy and values of a particular set of NPM reforms. On the one hand it could be said that such studies contain no empirical data on concrete impacts, and should thus be excluded. On the other it could be argued that this type of study can identify certain likely or logically necessary consequences of NPM, and therefore deserved to be included. Some studies of this type were in fact submitted to the database, but we have excluded them from most of our analyses, on the grounds that the discussion of ‘impacts’ really does require some empirical data and cannot just rest on *a priori* conceptual analysis.

How did we handle these inconsistencies? Basically, at the first stage we allowed all papers put forward by our COCOPS partners to be entered into the SharePoint database. However, subsequently we ourselves reviewed almost all the studies and were able to exclude some from certain sections of the meta-analysis (see next section). We also checked the coding of the metadata for inconsistencies, and eliminated these as far as we were able. Overall, therefore, we were fairly liberal in admitting studies to the database (out of a desire not to miss anything important) but rather strict in our subsequent interpretation of what these studies actually showed.

## **6. Meta-analysis**

COCOPS is committed to produce a meta-analysis, but what is a meta-analysis? Again, there is no single view of this, and definitions vary considerably.

Probably the dominant conceptualization is that prevalent in healthcare policy and medicine. Here a meta-analysis is a very strict set of statistical procedures by which the results of many different primary studies (typically randomized clinical trials) are compared so as to yield an overall average estimate of the net effect of the drug/procedure/treatment (Cooper, 1998). This is absolutely *not* the type of meta-analysis we have attempted – and with the kind of data that is available in the public management literature it would in any case be impossible.

An intermediate type would be still to attempt a statistical averaging of results, but to admit studies that were not experimental, such as an ordinary survey of managers' opinions, or of citizens' perceptions of changes in service quality. This is feasible and, indeed, has been attempted (Andrews et al., 2011). This approach to meta-analysis yields interesting results but it necessarily restricts itself to a certain type of study which, until now at least, has represented only a small fraction of the many and various writings about NPM and its effects. Thus one achieves a relatively precise answer, but on the basis of excluding the majority of both academic and practitioner studies. This did not recommend itself to us, at least partly because studies of the requisite type have been largely confined to the USA and the UK (see Andrews et al., 2011, Table A.1) and therefore we would not only have been drawing on just one small part of the literature, we would also have been excluding many EU member states entirely.

Our version is therefore more relaxed still. It includes some studies of the type that Andrews et al., 2011 reviewed, but many other types besides. It aspires to be what Pawson has termed 'realist synthesis', although at times it may be closer to 'narrative review' (Pawson, 2002; Pawson et al., 2005). Although different in many fundamental ways, classic meta-analysis, realist synthesis and narrative review all share a basic common aim:

'Meta-analysis is used to summarize and compare the results of studies produced by other researchers. A meta-analysis is often done to establish the state of research findings in a subject: in this way it provides the researcher with an overview of what others are saying about the subject rather than another discussion of one or a few parts of the question, problem or issue' (McNabb, 2010, p. 79)

One rule of classic meta-analysis which we have broken is the prescription that says 'only include studies which are based on original empirical research'. Whilst most of the studies in our database are based (at least in part) on original empirical research we have also allowed in

some which are more in the nature of synthetic, analytic overviews, i.e. analytic summaries of the work of others. The main reason is that these have sometimes contained useful interpretive insights which were not necessarily present in the original constituent studies that make up the synthesis. A lesser reason is that, in the public administration literature, both academic and even more practitioner, the line between original empirical research and secondary analysis is often far from clear. Many useful documents comprise an element of original research combined integrated with an analytic overview of other work. We saw no need to exclude this type of study from the database. Once again, we were fairly liberal in what we allowed in, but stricter in how we interpreted them.

In classic meta-analysis the intervention or programme is itself assumed to have causal powers. The hope is to identify the most powerful type of programme or intervention (Pawson, 2002, p. 341). But in the realist synthesis which we favour:

‘It is not programmes that work: rather it is the underlying reasons or resources that they offer subjects that generate change. Whether the choices or capacities on offer in an initiative are acted upon depends on the nature of their subjects and the circumstances of the initiative. The vital ingredients of programme *ontology* are thus its “generative mechanisms” and its “contiguous context”. Data extraction in a realist synthesis thus takes the form of an interrogation of the baseline inquiries for information on “what works for whom in what circumstances” ’ (Pawson, 2002, p. 342)

To put it simply, it is the *combination* of the *intervention* (policy or technique) and *context* which gives rise to specific impacts or results. So a given technique may produce quite different outputs and outcomes in different circumstances. For example, competitively contracting out refuse collection may, in a given context, turn out to be very successful, but that does not at all mean that the results of competitively contracting out brain surgery will be similarly positive. Contexts consist of complex assemblies of different elements, including interpersonal relations, organizational cultures, structures and procedures, legal frameworks and the political climate (Pawson et al., 2005). Because of our special interest in contexts we included a specific question in the meta-data section of SharePoint seeking to establish whether each study had identified specific contextual features and, if so, what they were.

## **7. Consequences of our definitions for the structure and scope of the study**

Some of our decisions increased the size of our databank, and some reduced it. Our relatively generous definition of what a meta-analysis could be increased our ‘stock’. If we had adopted

a strict medical style definition, in which we had only included randomized clinical trials (RCTs) of tightly defined reform-interventions, our database would have disappeared altogether! There is virtually no material of this kind with respect to major NPM reforms, in any EU country. Yes, there are some good outcome studies (although not that many, as we shall see later) but these are hardly ever based on the RCT model which is common in medical research.

On the other hand our definition of impacts excluded a great deal of material, both academic and practitioner, in which NPM is discussed at length, and various claims about it are made, but no real evidence of specific effects on processes, outputs or outcomes is presented. Thus, for example, there is a fair amount of both French and English material which discusses (sometimes with considerable sophistication and elegance) the philosophical and ethical implications of NPM thinking, but which includes no evidence of actual impacts. Similarly, there are spirited defences and attacks on NPM and ‘managerialism’ in the Anglo-American literature which score telling analytical points by relating it to broader neo-liberal doctrines or to a limited form of ‘technical rationality’, yet these accounts say nothing about anything concrete, and they do not advance any empirical evidence concerning impacts. We are not, of course, saying that these kinds of analysis are without value: we are simply saying that our focus is on impacts, and so they fall outside our terms of reference and, *ceteris paribus*, are not included in our database.

The results of our decision to include studies of processes (activities) are certainly debatable. This inclusion has considerably enlarged the range of material we are working with. And there is no doubt that some of the impacts – or claimed impacts – on internal organizational and political processes/activities are both interesting and important. For example, a number of studies claim that the creation, in the UK, of more than 130 executive agencies under the ‘Next Steps’ programme from 1988 resulted in a loss of capacity for policy co-ordination by ministries – that is a process impact (e.g. Office of Public Service Reform, 2002). On the other hand it has to be said that this finding, interesting though it undoubtedly is to ministers, civil servants and organization theorists, is still quite a long way from what the average citizen or media commentator might regard as an outcome. An outcome would be an event or condition *resulting* from the decreased policy co-ordination – for example, if people had not been able to get needed hospital care and had therefore died because hospital beds were being needlessly occupied by recovering patients who could not be discharged because co-ordination between health organizations and social care organizations had broken down. That would certainly be an impact *beyond* the organizations concerned, and we would classify it as an outcome.

## **8. Some preliminary remarks concerning the overall characteristics of the materials**

As we worked through the hundreds of documents in our database certain broad dimensions became apparent. Since these in an important sense ‘frame’ or condition the quantitative analysis provided in the next section it seems worthwhile setting them out here.

First, the proportion of documents that actually provide primary data concerning the *outcomes* of NPM-type reforms is very small. Such articles (e.g. Propper et al., 2008; Kelman and Friedman, 2009) are very valuable but also rare. It is probably no accident that these gems tend to occur most often in the historically two most professionalized and scientifically-based public services – healthcare and education. They also tend to come from those countries that maintain large performance databases, especially the UK and (beyond this study) the USA. There are many more documents that look at *outputs*, but even these are not in a majority. At the other extreme, the number of secondary studies which discuss NPM reforms in a general way, citing other literature, but with no original data on outputs or outcomes is very large. (There are several reasons why this should be so, as indicated in section 4, above.) So, it might be said, the sheer size of the NPM literature is potentially misleading – only a small part of it directly pertains to our key question: what impacts has NPM had on outputs and outcomes? Or, to put the same observation in a more critical way, lots of people have had lots of ideas about NPM (many of them very interesting and persuasive) but rather few have gone out into the field to collect primary evidence concerning specific impacts on service users and the citizenry more generally.

Second, it is clear that the difficulties we have had in deciding on, and then keeping to, a consistent definition of activities/processes, outputs and outcomes are difficulties which are far from being ours alone. In both official reports and the academic literature the line between processes and outputs (in particular) has often been treated in a very plastic way. Thus, for example, producing a report by a target date is often treated as an ‘output’, although we would normally see that as part of the internal process of government (few citizens are going to read or even hear of this report). Or again, achieving particular reorganizations are frequently treated as outputs or even outcomes – ‘we set up two new agencies and required them to operate to strict performance targets’ – this is interesting information, but to us it denotes process and activity, not output, and certainly not an outcome. The existence of the agencies and their targets in no way guarantees that the citizens and businesses using their services will have experienced any change in the level or quality of what they receive. It is significant that one of the best collections of comparative data on government reforms yet produced, the OECD’s *Government at a Glance* (2009) is principally concerned with (in our

terms) inputs and processes. More than one hundred and fifty pages of data do not contain a single table showing an input/output ratio (efficiency), still less a set of attributable outcomes (although the most recent, 2011 edition does tentatively begin some input/output analysis). Even when academics undertake extensive fieldwork in search of information about how reforms have affected performance, they may be unable to find it (Pollitt et al., 1998a, pp.162-166). However, many academic articles describe how far techniques such as performance management or Total Quality Management have been installed in specified public departments or agencies (e.g. Joss and Kogan, 1995; Lægheid et al., 2008; Torres and Pina, 2004; Verheijen and Dobrolyubova, 2007). For us, once more, this is interesting, but not yet an output, still less an outcome. *If* the installation of TQM can be shown to have shortened public waiting times, then that *is* an output. And if the shortened waiting times and earlier treatment can be shown to have improved the outcomes for specific health problems (e.g. in cancer treatment or addiction problems) then those would be outcomes. However we should not assume that the mere articulation of a set of procedures (in this case TQM) will itself lead to such changes.

Thus we have taken a relatively relaxed approach to the process of admitting documents to our database, but a slightly puritanical approach to categorizing and analyzing what these documents finally contain. We make no apology for this latter severity. Public management reform is a field extensively populated with individuals and organizations who have a stake in claiming successes. It is also a field, as we have seen, where the definition of many of the key terms is slippery and various. Our puritanism, if that is what it is, stems from a desire to achieve a measure of consistency across a large number of very diverse documents, together with our determination to focus, as far as we are able, on the outputs and outcomes that matter to the world *outside* public sector organizations. The question, to put it crudely, is ‘what do we *get* for all this reorganization?’ where ‘we’ are the citizenry and its representatives, civil society associations and companies. In all the multifarious literature on NPM this question has not been asked often enough, and has rarely been asked on the empirical scale that is attempted here in COCOPS.

## **9. Key descriptive statistics from the database**

Having now travelled through this dense thicket of definitional problems and methodological choices we can now begin to look at what has actually been found. According to the selection criteria described above, we ended up with a grand total of *518 documents*. These were distributed between the different countries and the European Commission as follows:

**Table 2. Distribution of studies by country**

<b>Country</b>	<b>Number of studies*</b>	<b>Percentage**</b>
Austria (AT)	29	5.6
Belgium (BE)	39	7.6
Croatia (HR)	20	3.9
Czech Republic (CZ)	25	4.9
Denmark (DK)	19	3.7
Estonia (ES)	24	4.7
Finland (FI)	15	2.9
France (FR)	71	13.8
Germany (DE)	64	12.4
Greece (GR)	12	2.3
Hungary (HU)	46	8.9
Ireland (IE)	13	2.5
Italy (IT)	60	11.7
Latvia (LV)	10	1.9
Lithuania (LT)	17	3.3
Luxembourg (LU)	10	1.9
Netherlands (NL)	65	12.4
Norway (NO)	38	7.4
Poland (PL)	13	2.5
Portugal (PT)	20	3.9
Romania (RO)	21	4.1
Slovakia (SK)	20	3.9
Spain (ES)	23	4.5
Sweden (SE)	26	5.0
Switzerland (CH)	15	2.9
United Kingdom (UK)	101	19.6
European Commission (EC)	8	1.6

\* Includes both single-country and comparative studies. To avoid double counting the studies should not be added across countries. \*\* Percentage of the total database of 518 studies.

There are some evident differences in the number of studies included for various countries with five of them, the UK, Germany, the Netherlands, France and Italy, being better represented (each more than 10% of the total studies) than the rest. Overall, however, for most states (18 out of 27) the difference in the numbers is small, each of them with fewer than 5% of the total database. The figures include both single country and comparative studies, with varying degrees of treatment; therefore simply because a country had more studies included in the database does not automatically mean that it received a more intense

treatment. Table 2 is complementary to table 1 in which we reported the degree of attention we had been able to give to each country.

Tables 3 to 6 below describe the database with respect to sectors, type of documents, methods used and particular NPM-type reforms, tools and mechanisms identified.

The database suggests that the extent of evaluation of NPM reforms varies across sectors. Some sectors, such as education and health, appear to have been subjected to more evaluation than others, such as defence or spatial and urban planning. This may also reflect different degrees to which NPM-type reforms have been implemented across sectors, with health and education in the lead. Alternatively it may mean that those sectors have stronger traditions of, and/or capacities for, such evaluation – i.e. that other sectors may have experienced equally or even more intense NPM initiatives, but have not subjected them to the same kind of organized scrutiny. It may be pertinent that healthcare and education both have considerable ‘attached’ communities of independent scholars, in a way that, say, defence or the issuing of driving licenses do not.

One can also notice that there were a sizable number of studies which evaluated parts of central government such as the core executive, the central government or the public sector more generally. Although our emphasis was on central government reforms, a significant number of local level studies were eventually included, as shown in table 3. This was because our COCOPS partners often put forward studies of local government as the best exemplars of analyses of NPM impacts. This was the case both in Western Europe and the CEE.

**Table 3. Distribution of studies by sector**

<b>Sector</b>	<b>Number of studies*</b>	<b>Percentage**</b>
Core executive (excluding line ministries)	138	26.6
Legislature and associated bodies	18	3.5
Employment services	73	14.2
Health care, education and social services, except employment	184	35.7
Business and economic policy	74	14.4
Defence	30	5.8
Justice, law and order	57	11.1
Spatial and urban planning	22	4.3
Local services/local government	110	21.4
Public sector more generally, including	151	29.2

central government		
Other	25	4.9

\* Due to many multiple-sector studies, the columns should not be added (to avoid double counting).

\*\* Percentage of the total database of 518 studies.

**Table 4. Distribution of studies by type**

Type of study	Number of studies	Percentage*
Academic	352	68.0
General official policy reports with some elements of evaluation	35	6.8
Internal formal official evaluation studies	29	5.6
External and independent official evaluation studies	59	11.5
Management consultancy reports	18	3.5
Studies by international or supranational bodies	13	2.5
Studies by civil society organizations	6	1.2
Other	7	1.4

\* Percentage of the total database of 518 studies.

Academic studies by far outnumber any of the other type of study. However, more than 30% of the total database consists of non-academic studies (table 4). We believe that going beyond conventional approaches (i.e. including only academic studies using one preferred type of method) can add data and value to our analysis and conclusions. In this sense our meta-analysis differs from other meta-analyses in academic public management. Furthermore, we would point out that some of the non-academic studies were at least as rigorous as many of academic ones (e.g. some performance audits by national audit offices). Indeed, the categories ‘academic’ and ‘official’ are far from mutually watertight, because some of the official studies are carried out by academics hired for the purpose, and some of the best of the ‘pure’ academic studies actually rely heavily on official databases.

**Table 5. Distribution of studies by methods**

Methods	Number of studies*	Percentage**
Mainly declarative (limited evidence)	87	16.9
Broad synthetic overview (there is an analytic attempt to make an assessment)	229	44.2
Single case study	67	12.9
Multiple case studies	119	22.9

Hypothesis/hypotheses testing	19	3.7
Historical descriptive narrative	45	8.7
Experimental method	4	0.8
Mainly quantitative analysis, but not experimental	108	21.0

\* Due to many multiple-method studies, the columns should not be added (to avoid double counting).

\*\* Percentage of the total database of 518 studies.

**Table 6. Distribution of studies by NPM reforms, tools and mechanisms**

<b>NPM tools and mechanisms</b>	<b>Number of studies*</b>	<b>Percentage**</b>
Corporatization	39	7.6
Market-type mechanisms more generally (MTMs)	114	21.9
Contracting out	73	14.2
Purchaser-provider split	18	3.5
Performance league tables	27	5.2
Performance-related budgeting	43	8.4
Performance-related pay	41	7.0
User-choice mechanisms	21	4.1
Other MTMs	21	4.1
Performance measurement and management systems more generally	218	42.1
Normalization of public sector employment	41	8.0
Employment of management consultants	12	2.3
Use of internal contractual/quasi-contractual frameworks for steering	82	15.9
Creation of semi-autonomous agencies	72	13.9
Public service quality improvement schemes	60	11.7

\* Due to many multiple-tool/mechanism studies, the columns should not be added (to avoid double counting).

\*\* Percentage of the total database of 518 studies.

Table 6 above suggests that the various NPM tools and mechanisms have not been evaluated to equal extents. Some reforms, such as performance management and measurement systems or contracting out, or market-type mechanisms appear in the literature more often than other type of reforms. [Our categories were designed to group together (and have grouped together) broadly similar tools and mechanisms.]

## **10. A preliminary interpretation of the findings**

Our various analyses of the database relate to different subsets of the database contents, and we always specify which subset is in play. In some cases (most importantly in section 11 on

contextual influences – see below) we have analysed only those studies which include findings about outputs and outcomes. In others we have also included studies which focus on changes in activities/processes and may say little or nothing about outputs and outcomes. Throughout we have followed the pragmatic line announced in section 3 above, in the sense that we have included studies of the local impacts of reforms (e.g. in schools and hospitals) so long as the reform itself was largely initiated by central government (for example, the study examined in section 12 is of this type). Table 7 (below) presents the major subsets we have used, in descending order of size.

**Table 7. Some subsets of the COCOPS SharePoint database**

<b>Subset</b>	<b>Number of studies</b>	<b>Percentage*</b>
1. All studies (total database)	518	100.0
2. All studies with at least one entry for effects in general**	436	84.2
3. All studies including entries for changes in activities/processes, outputs or outcomes	354	68.3
4. All studies including entries for changes in outputs or outcomes	138	26.6
5. All studies with entries for changes in outputs or outcomes in central government only	117	22.6
6. All studies with entries for changes in outcomes	45	8.7
7. All studies with entries for changes in outcomes in central government only	39	7.6
8. All studies with entries for changes in quality	61	11.8
9. All studies with entries for changes in social cohesion	135	26.2

\* Percentage of the total database of 518 studies.

\*\* We coded a variety of effects that the studies in our database identified. These included, for instance, effects on transparency, accountability or organizational stability and fragmentation of the public sector. However, as indicated earlier, in this paper we have treated changes in processes/activities, outputs and outcomes (row 3) as impacts, and have devoted specific attention to changes in outputs and outcomes (rows 4-7).

One could say that the further one goes down the list from 1 to 7, the closer one comes to our ideal aim – to establish what actual final impacts (outcomes) NPM reforms in central

government have produced. Rows 8 and 9 (quality and social cohesion) are somewhat separate and complex dimensions, as indicated by the discussion previously in this paper.

A total of 518 studies and reports of various types have been reviewed and coded according to an analytical framework. In 436 (second row in Table 7) of them at least one type of effect (generously construed) has been identified (see footnote \*\* in Table 7 above). Many of the studies we reviewed are complex and contain various types of effect, from, say, transparency and accountability to changes in staff motivation and morale to outputs and outcomes.

One clear conclusion from table 7 is that many of the studies have not gone beyond reporting changes (or no change) in processes or activities. Effects on outputs are less common than changes in processes/activities and they typically include effects on efficiency and productivity, quality of services and quantity of outputs. As for outcomes, we found that only 45 studies of the total database (less than one out of ten) have gone that far. A typical outcomes-related measure is effectiveness, or the degree to which outcomes meet stated objectives of the reforms. One can adopt an even more conservative definition of effectiveness by including only measures of effectiveness that pertain to effects produced in the real, outside world of users and citizens or businesses. In most of the 45 studies this has been the case, but in some cases we allowed a more generous definition of effectiveness and treated it as an outcome in itself by its virtue of reflecting goals that have been accomplished. Subtracting the studies that evaluated outcomes of local services from the total of outcome studies, we ended up with only 39 or 7.6% of the total database with at least one entry for outcomes in central government.

Finally, 61 studies in the database had at least one entry for quality, mostly quality of services or quality in operations, processes or activities. Based on our categorization of social cohesion in terms of effects on equal access to services and on civil servants' motivation and morale, we found that 135 studies included at least some comment on social cohesion.

Below in tables 8 and 9 we report additional findings. The data seems to suggest that overall the impacts of NPM reforms have been more often assessed as positive than negative. One should note, however, that this broadly and apparently favourable picture is by no means uniformly positive. As table 8 shows, we found many studies with some evidence or comments concerning negative changes in outcomes, outputs or processes/activities. A sizable number of studies have not reported significant changes (or have reported no changes at all). More than 30% of the total entries for outcomes are of this kind, 27% for outputs and

more than 23% for processes/activities. That makes *prima facie* sense, in that it is generally less difficult to change processes than outputs, and outputs than final outcomes.

**Table 8. Summary of impacts of NPM-type reforms in Europe**

Direction  Extent of impacts	Up/Improved		Down/Worse		Unchanged or uncertain		Total number of entries**
	Number of entries	%*	Number of entries	%	Number of entries	%	
Outcomes	25	43.9	13	22.8	19	33.3	<b>57</b>
Outputs	87	53.4	32	19.6	46	27.9	<b>165</b>
Processes/ activities	373	57.9	119	18.5	152	23.6	<b>644</b>

\* For each type of impact the percentages are based on the total number of entries for that particular type of impact.

\*\*Some studies have contained one or more entries for outcomes, outputs or processes/activities. Some have also included more than one entry of the same type of impact, i.e. outcomes, outputs or processes/activities. Therefore an entry is not the same as a study. The total number of entries column shows the sum of all entries for each particular type of impact (namely for outcomes, outputs and processes/activities).

Another point to note when interpreting the findings is that the labels we used, such as ‘up’ or ‘improved’ or ‘worse’ can carry either qualitative or quantitative connotations, or both. This is particularly so for effects on processes/activities and to some extent for outputs, but less so for outcomes which, by their very essence, are qualitative (‘soft’ impacts) *par excellence*. In reporting the findings we have not distinguished between qualitative improvements in processes and ‘simple’ changes in processes, such as the introduction of a performance management system or a public service quality improvement scheme. The latter may or may not lead to positive improvements in the quality of the process, and therefore improvements in outputs and possibly outcomes. Most of the 373 entries for positive changes in processes reflect quantitative changes in systems, processes and activities (i.e. they record that something new has been introduced) and only a minority reflects documented qualitative changes (i.e. the process has actually improved, against some relevant normative standard).

What is the story behind the recorded impacts on outputs and outcomes? Table 9 below presents more specific findings on the identified outputs (not outcomes). The limited number of studies that have attempted and/or succeeded in assessing outcomes makes it difficult to conduct more specific analyses of changes in outcomes. However, later in the paper we refer in more detail to studies that have found such changes.

**Table 9. Summary of findings on outputs**

<b>Direction</b>  <b>'Type' of outputs</b>	<b>Up/Improved</b>		<b>Down/Worse</b>		<b>Unchanged or uncertain</b>		<b>Total number of entries**</b>
	Number of entries	%*	Number of entries	%	Number of entries	%	
Quantity of outputs and services	20	71.4	4	14.3	4	14.3	<b>28</b>
Quality of outputs and services	16	43.2	10	27.0	12	31.6	<b>38</b>
Efficiency and productivity	39	54.2	11	15.3	23	31.5	<b>73</b>
Other	12	42.9	7	25.0	9	32.1	<b>28</b>

\* For each type of output the percentages are based on the total number of entries for that particular category.

\*\*As in the previous table, some studies have included more than one entry for outputs, therefore an entry is not the same as a study. The total number of entries column shows the sum of all entries for each particular type of output.

The data on outputs suggest that:

- More studies appear to have assessed the effects of NPM-type reforms on efficiency and productivity than on changes in quality and quantity of outputs (73 entries compared to 38 and 28 respectively). However, to measure productivity or efficiency we require data that relates inputs and outputs, i.e. a measure of the changing ratio between them, but few studies have succeeded in doing so (see comment below)
- Measures related to outputs regardless of their type appear to have improved in more cases than they were found to have decreased or remained unchanged
- This picture is particularly evident in the case of quantity of outputs and less so for quality of outputs and for efficiency and productivity
- A more significant number of entries on quality and efficiency or productivity (close to 30%) compared to those on quantity of outputs were found, which may show the complex nature of assessing these impacts.

However, as has been emphasized earlier in this paper, a number of challenges make the evaluation of the impacts of NPM reforms problematic. In the case of outputs, the loose use of terms efficiency, productivity or quality and the lack of evidence supporting such claims are especially salient. Many studies have made claims that efficiency has changed, but have failed to provide evidence of a ratio between inputs and outputs. More often than not, these claims are based on findings indicating that systems work faster or better and activities have improved, but little or no data on the costs/budgets or inputs is considered. Inputs may have

well decreased, stayed the same or increased during this time. Conversely, we have often found claims on changes in efficiency with little or no evidence on the output side of the ratio. In fact these are referring to changes in inputs, savings or economies, which are not efficiency measures (e.g., Centraal Planbureau, 2007; Dankó and Molnár, 2009; Office of Public Services Reform, 2002; Rossmann, 2001). A more detailed example of this common challenge will be given in section 12 below. Another point is that the costs of evaluating reforms are seldom included in the evaluation itself – the latter being considered as something separate. Yet from a broader perspective it could be argued that a proper evaluation is an integral part of management reform, and its cost should therefore be included as one component of the transaction costs of the reform.

A further point is that improvements in one aspect may be balanced out by deteriorations in another. Thus, to say that the majority of those studies which focus on outputs record improvements does not mean that there may not also be deteriorations recorded in some of those studies as well. For example, efficiency may improve but equity of access declines. Or quality of service is improved but only at the cost of higher input expenditure.

Overall, therefore, one may argue that the glass is half full, or half empty. NPM enthusiasts can point out (Table 8) that ‘up/improved’ is the largest of our three categories of impact for outcomes, outputs and processes/activities. However, NPM sceptics can equally make the very serious point that, as far as outcomes are concerned, more than half the entries in this large database show either no change or worse outcomes after NPM reforms. The temptation to draw easy, general conclusions should be resisted – taken together, what Tables 8 and 9 most obviously demonstrate is the huge variation in the impacts of NPM reforms. NPM cannot be regarded as some kind of ‘silver bullet’ which regularly and reliably produces improvements in either outputs or outcomes.

## **11. Contextual influences**

Contextual influences can mean that even similar-sounding reforms in one country or sector can vary considerably from one organization to another (Joss and Kogan, 1995). Contextual variation becomes even more obvious in international comparisons, where some national features may wholly undermine – or even prevent the emergence – of a particular type of reform (e.g. Wilson, 2011). Many of the studies in the database mention contextual influences. These can evidently affect each kind of impact – activities, outputs and outcomes. Thus our finding that contextual factors are important in Europe is simply in line with a large body of public management scholarship that has stressed the importance of these influences. Particular reforms fare very differently depending on where, when and with whom they are

attempted (see, e.g. Lynn, 2006; Ongaro, 2009; Painter and Peters, 2010; Pollitt et al., 2004; Pollitt and Bouckaert, 2011).

But what are ‘contextual factors’ and how do they work? Many authors mention contextual influences, but few seem inclined to classify or theorize them. In fact this is a very complex issue, and hard to conceptualise, because the list of possible contextual influences is very varied and very long. One (widely cited) approach is to identify contextual variation with local power structures, for example:

‘We know what rationality is, and rationality is supposed to be constant over time and place. This study, however, reveals that rationality is context-dependent and that the context of rationality is power. Power blurs the dividing line between rationality and rationalization.’ (Flyvberg, 1998, p. 2)

We have some sympathy with this view. Certainly there are occasions when public management reforms are pushed forward by powerful politicians or officials who ignore rational drawbacks and, in effect, ‘rationalize them away’. However, we also see advantages in taking a broader conception of contextuality than Flyvberg. From our perspective it is not always just about the exercise of power. Contextual influences can also include chance happenings or unintentional mistakes and miscalculations. They can include institutional structures which, though they may represent some fossilized power bargain long ago, do not directly reflect the aspirations of power-holders today, and may even inhibit those aspirations. So power is very important, but it is not the whole of contextuality.

One useful approach has been outlined by Löffler (2000). She categorised what we would term contextual influences into three broad types of ‘contingency’:

*Stimuli.* Events that occur at the initial stage of an innovation/modernization process and create a critical situation that induces political and administrative actors to search for or accept new solutions

*Structural variables.* Institutional and cultural characteristics of a given administrative and political system that are necessary conditions for a specific best practice to work

*Implementation barriers.* Institutional and cultural characteristics of the administrative and political system and behaviours of stakeholders that inhibit the successful implementation of foreign best practice cases in the domestic context’ (Löffler, 2000, p. 201).

It is not difficult to see examples of all three types in our database. A case of stimulus would be the ‘Kinnock reforms’ in the EU Commission (2000-2004) where the collapse of the

previous Santer commission in disgrace gave a huge stimulus to rapid reform, of a systemic type that the Commission had hitherto resisted or avoided for decades (Bauer, 2009; Pollitt and Bouckaert, 2011, pp. 256-261). Another would be where well-publicized cases of failure in hospitals put governments in a position to install tighter regimes of performance measurement and management control of professionals (Pollitt et al., 2010). A case of structural variables having an influence would be the way in which so many of the management reforms in the CEE countries 1995-2010 were of modest effect partly because of the limited capacity of CEE administrations (Meyer-Sahling, 2009; Verheijen, 2007). How were CEE governments supposed to contract out on a large scale when they had so few civil servants with any training or skill in contract management (Lember, 2008)? Examples of implementation barriers are also not hard to find. Torres and Pina (2004) argue that the political culture in Spain inhibited the shift towards a more NPM-like Human Resource Management system, while in the UK no such barrier impeded reform. Meyer-Sahling (2009) shows that the rapid turnover of governments in a number of CEE countries militates against longer-term, strategic types of reform (e.g. in HRM) where sustained support is needed over a period of years.

Useful though the Löffler threefold categorisation is, however, we think that for the purposes of COCOPS, we need a slightly more refined framework for contextual influences. We have developed this inductively, by following the following steps:

1. We read systematically through all the studies in the database which identified specific changes in outputs or outcomes
2. We noted wherever one of these studies mentioned a contextual factor as having been an important influence, and listed what those factors were
3. We inspected the list and built from it a set of categories that are appropriate to our specific topic – the impacts of NPM reforms.

Three particularly important features emerged from this exercise. First, one can conceptualize a range from deep-seated, long-lasting influences at one pole towards medium term and then short term influences at the other. On the whole reformers should be able to foresee and anticipate many of the longer lasting influences (e.g. those introducing reforms in the CEE countries in the 1990s should have recognised that they were starting with a civil service that was weak on many skills, often corrupt and not at all ‘consumer focused’ – and they should therefore have included plans for dealing with these features). But the closer one comes to the other pole – sudden events and chance happenings – the more difficult it will usually be for reformers to plan or allow for them (e.g. reforms may be blown off course by the discovery of a major scandal, or the occurrence of a major accident or natural disaster just after the reform

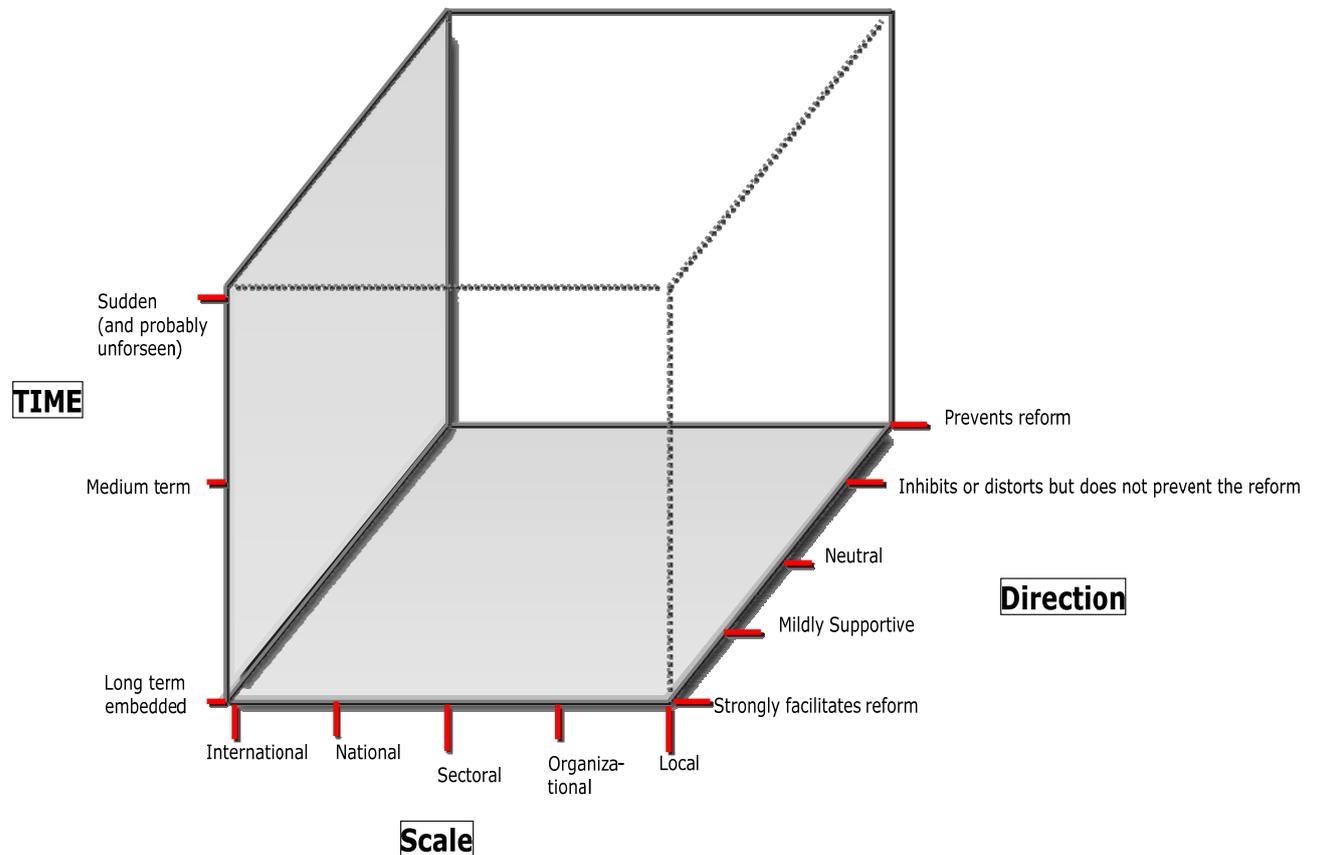
has been announced). There is a complication here which needs to be noted. Time runs in both directions. Some sudden, surprise events can have long term consequences in the future (9/11 being an outstanding example). Other contextual factors may have existed for a long time but have only a minor influence in the future. On the whole, however, features of political systems and public organizations which have existed for a long time – cultures, basic structures, embedded standing operating procedures – are likely to be resistant to being changed overnight. In short, it is highly probable that long-standing, basic features of the machine will require sustained efforts over months or years before they can be removed or radically modified.

Second, different contextual effects are not equally broad in their effects. Thus the civil service culture may be expected to exert an influence right across government. Equally the effects of a centralized, powerful executive are likely to be widely spread. However not all contextual influences are on this scale. Some may be local, yet nevertheless durable and strong. For example, it may be that a particular town or a particularly prestigious organization has a different culture from the others and insists on ‘doing things differently’ (e.g. the case of the capital, Tallinn, in the primary health care reforms in Estonia – see Atun et al., 2005). Or there may be a local incident or controversy, such as that concerning the shooting incident which led to the premature departure of a reforming English police chief, as cited in Pollitt and Bouckaert, 2009 (pp. 111 and 150-151).

The third important feature is that contextual influences can help or hinder reforms. Indeed, the politico-administrative culture, for example, may hinder one type of reform but help another. A political system with a powerful centralized executive may help that country quickly to implement a radical reform that improves, say, tax collection services, but on another occasion the same structural features may encourage governments to over-reform – to go on and on changing things just because they are easy to change, until confusion and cynicism begin to grow (Pollitt, 2007; Pollitt and Bouckaert, 2011). Thus, although contextual factors most frequently gain attention when they hinder reforms, they can also facilitate implementation.

Figure 2 below offers a graphic schema which incorporates each of these three points. It offers what we hope is a somewhat more advanced (or at least more differentiated) conceptual framework for thinking about contextual influences than the earlier one from Löffler (see above).

**Figure 2: A conceptual framework for classifying contextual influences on public management reforms**



This framework can be used to ‘locate’ the kinds of contextual influences which the authors of many of our studies identify. Thus, for example, the socio-economic status of the local catchment area tends strongly to affect the exam results achieved by a school (Wilson and Piebalga, 2008). What kind of local contextual influence has, say, a deprived socio-economic catchment on a national reform designed to improve exam performance? In terms of Figure 2 it is:

- Local in scale (because the socio-economic composition may be quite different a kilometre or two away)
- Medium to long term in time (the socio-economic composition of neighbourhoods can and do change over time, but that is usually a matter of years or decades rather than weeks or months – and it cannot be controlled by the school)
- Inhibiting or distorting of reform (it does not stop the reform being implemented, but it probably reduces its effectiveness in comparison with a school that has a catchment in a locality with a high socio-economic character).

To take another example, if we look at a study of the marketization of healthcare services in Central and Eastern Europe (Nemec and Kolisnichenko, 2006) we find that the (largely) disappointing achievements of reform were deeply affected by wider economic conditions at the time.

‘probably the crucial objective issue in all countries was finance, which was tied to the level of economic performance of the country. Health reforms started to be realized in the period of more or less massive decline of GDP per capita in most CEE countries’ (Nemec and Kolisnichenko, 2006, p. 24).

In terms of Figure 2 this would be:

- International in scale (it affected the whole CEE region)
- Somewhere between ‘quite sudden’ and ‘medium term’ in time (the decline in GDP per capita began quite suddenly but lasted for a few years – reformers could hardly have claimed to be unaware of it)
- Inhibiting or distorting of reform (the reforms went ahead but did not work out nearly as positively as had been proclaimed)

Having established this framework we can now return to the database and illustrate some examples of the kinds of contextual influences which it yields. These are shown in Table 10 (below).

**Table 10: Contextual influences on public management reforms**

<i>Contextual influence</i>	<i>Scale of influence</i>	<i>Time needed to achieve change</i>	<i>Direction (supports or inhibits NPM)</i>	<i>Examples</i>	<i>Comments</i>
<i>1. Politico-administrative culture</i>	National, sectoral	Medium to long term	Studies of countries with a <i>Rechtsstaat</i> administrative tradition have typically found politico-administrative culture to inhibit NPM-type reform. Post-communist cultures (CEE), for example, may allow the reform to take place but subvert its implementation	Brunetta, 2009; Capano, 2003; Læg Reid et al., 2008; Nemeč, 2007; OECD, 2010; Pollitt and Bouckaert, 2011; Societatea Academică Română, 2010; Șandor and Tripon, 2008; Torres and Pina, 2004	The politico-administrative culture in a number of countries in CEE and elsewhere in Europe (e.g., in Italy or Spain) has often been characterized by excessive legalism, low citizen trust and confidence in the public sector, poor satisfaction with civil servants, lack of transparency and issues of patronage and corruption.
<i>2. Structure of the political system (e.g., centralized/decentralized; majoritarian/consensual)</i>	National	Medium to long term	Supportive or inhibiting of NPM-type reforms, depending on the characteristics of the political system	Askim et al., 2010; Christensen, 2001; Christensen et al., 2006; Minvielle, 2006; Norges Forskningsråd, 2007; Pollitt et al., 2004; Pollitt and Bouckaert, 2011; Roberts, 1997	Major differences exist between the incremental decision-making process (in e.g., Germany or Norway) compared to the majoritarian and centralized political system in the UK. The latter facilitates the implementation of rapid and large-scale reforms.

3. <i>Rapid turnover of governments and/or rapid change in governmental structures and functions and ministerial appointments</i>	National, sectoral, ministry-level or organizational	This can be short term, but more often it is a medium term pattern of government instability	Tends to reduce the continuity and coordination of reform efforts in time or across government	Dunleavy and Carrera, 2011; Meyer-Sahling, 2009; Şandor and Tripon, 2008; Verheijen, 2007, p. xvi	A particularly common factor in CEE, but not exclusively a CEE problem (see e.g., Dunleavy and Carrera for an example of the UK Department of Work and Pensions).
4. <i>Lack of current administrative capacity (e.g., weak or inadequate contract management skills, strategic planning skills or leadership skills)</i>	National, sectoral or organizational	Medium term – a long-standing skill deficit will take resources and typically years rather than months to remove	Potentially severely inhibiting of reform. May mean that reform cannot be ‘rolled out’ but can only be implemented in a few specially resourced and staffed model projects	Atun et al., 2005; Baggott, 1997; Edelenbos and Klijn, 2009; Hammerschmid et al., 2011; Kuhlmann et al., 2008; Laforgue, 2004; Lember, 2008; Macinati, 2006; Meyer-Sahling, 2009; National Audit Office, 2009; Nemec, 2008; Promberger, 2008; State Audit Office, 2002; Randma-Liiv, 2005; Verheijen, 2007; Verhoest, 2005; Yesilkagit and De Vries, 2002	Common in CEE countries. However, this factor was also found to be important in a number of countries in the West, including the UK and Germany. Experience of working closely with the private sector and the possession of leadership skills have been found to be positively associated with good performance.
5. <i>Socio-demographic</i>	Regional or	Usually medium	Particular user characteristics may	Andersen, 2008;	The performance of students in

<i>characteristics of geographical areas, service users or citizens</i>	local	to long term. The socio-demographic characteristics of particular localities do change, but usually not overnight.	help or hinder NPM-type reforms. In education, for example, simple versions of parent choice tend to advantage the more mobile, better-informed middle class families	Atun et al., 2005; Wilson and Piebalga, 2008	the Danish public school system, for instance, has been influenced more by their socio-economic status than by the introduction of performance measurement (Andersen, 2008).
<i>6. Need for non-standard individual treatments of service delivery (i.e. the task itself is complex and variable)</i>	Organizational and sometimes sectoral	In so far as this need is embedded in the structure of the task, it may well be difficult to change in the short term. It sometimes alters with technological change in the longer term.	Typically at odds with the NPM focus on standardization in order to achieve efficiency and quality control. Potentially leading to major distortions and user dissatisfaction with the service	Belorgey, 2010; Divay, 2009; Harrison and Pollitt, 1994; Matelly and Mouhanna, 2007; Pollitt at al., 1998b; Pollitt et al., 2004	Particularly relevant to social and human services. The NPM focus on rapidity and efficiency in service delivery was found to negatively affect users' satisfaction with the service in sectors such as employment services (Divay, 2009), hospitals (Belorgey, 2010) and police (Matelly and Mouhanna, 2007). Problems include professional resistance to loss of discretion, plus possible gaming of performance measurement systems. This is a particularly salient factor in the French NPM literature, but also occurs elsewhere.
<i>7. International (external) pressure for reform</i>	International (though it may be focused on	Short to medium term	Supportive of reform – though not necessarily the most appropriate types of reform. Relaxation of	Arndt, 2008; Meyer-Sahling, 2009; Nemeč and	Pressure from the World Bank and the IMF, for instance, for market-type healthcare reform

	specific organizations or sectors)		international pressure may result in a weakening of the reform programme (Meyer-Sahling, 2009)	Kolisnichenko, 2006; Pollitt, 2010	in a number of CEE countries proved influential but misplaced since these countries were unprepared for such reforms. The EU accession efforts in Romania and in other CEE countries are often said to be a main driver, accepted by parties across the political spectrum. This helped to improve the continuity of reform efforts in the pre-accession period, despite government changes.
8. <i>Organized resistance from stakeholders</i>	National, sectoral or local	Short to medium term	Inhibiting or possibly preventing NPM-type reforms	For professional resistance see Belorgey, 2010; Divay, 2009; Harrison and Pollitt, 1994; Matelly and Mouhanna, 2007; Pollitt at al., 1998b; Pollitt et al., 2004. For union resistance see Capano, 2003; Emery and Giauque, 2001; Hondeghe and Depré, 2005	Highly qualified professionals typically resist the NPM focus on standardization. Opposition from labor unions was a key factor that inhibited and even prevented the full implementation of the Copernicus reform in Belgium (Hondeghe and Depré, 2005).

9. <i>Sudden accidents or scandals</i>	Organizational	Short term in the sense that they are unpredictable, but may have longer term consequences if rules and structures are changed	Typically stimulating reform, although a major scandal or disaster may inhibit or even prevent reform	Bauer, 2009; Pollitt and Bouckaert, 2011; Pollitt et al., 2010	For example, the scandal preceding the Kinnock reforms in the EU Commission stimulated the decision to implement comprehensive reform.
10. <i>Contextual features influencing performance targets and other performance incentives</i>	National, sectoral, organizational	Short to medium term	Performance management schemes are likely to yield negative results and be unsustainable if poorly defined and/or implemented in a highly politicized context	Andersen and Pallesen, 2008; Behaghel et al., 2009; Kelman and Friedman, 2009; Nemeč, 2007; Pollitt et al., 2010; Propper et al., 2008	Particularly relevant to health and education, but also to employment services. Examples have been identified in a number of countries including the UK, the Netherlands, France, Denmark and Slovakia.
11. <i>Contextual features affecting performance-related pay (PRP)</i>	National, sectoral, departmental	Short to medium term	PRP only works well when perceived as supportive, motivational, fair and clear. If otherwise it can lead to negative consequences, such as lack of cohesion, decrease in staff collaboration and lowered morale	Emery, 2004; Jeannot and Guillemot, 2010; Makinson, 2000; Marsden and French, 1998; Randma-Liiv, 2005; World Bank, 2001	The existence or lack thereof of managerial capacity has been found to be a critical factor. Likewise, political patronage, favoritism and corruption, small bonuses, the difficulty of measuring results objectively and the lack of legitimacy of the scheme have been found to negatively affect the effectiveness and sustainability of PRP schemes.

Table 10 is complex enough to need some explanation. A first point to note is the wide variety of contextual factors that have been found to have an important influence on the impacts of NPM reforms. They range from deep-seated, embedded influences such as the politico-administrative culture and structure of political system to short-term, but influential factors such as the role of leadership or a sudden scandal or accident.

In table 10 we have condensed a large number of accounts into 11 categories of contextual variables. Probably this process could usefully go further, so that greater simplicity and taxonomic elegance would be achieved. At our present stage of understanding, however, we have refrained from doing that, preferring to stay reasonably close to the designations given by the authors in our database.

There is an order or sequence within the table, although it is not absolutely precise. Basically the table begins with deeply embedded contextual features which are likely to be very hard to transform in the short term. These are therefore contextual elements which, at least at the beginning, reformers must learn to work with or through rather than sweep out of the way or ignore. The table then progresses through elements which can in principle be changed in the medium term and ends with those which may change suddenly or at short notice (which is not, of course, the same as saying they can be easily controlled). This rough order or sequence can be seen in the first nine rows (1 to 9).

Somewhat separately from this sequence we have also identified a number of studies of two particular NPM component elements – performance management in general (indicators, targets, incentives) and performance-related pay in particular (PRP). Each of these components is reported in our database studies as being heavily influenced by particular sets of contextual influences. These are listed at the end of the table (categories 10 and 11) and can be thought of as mixtures of longer term and shorter term factors which influence these two particular components of NPM (first, performance management generally and secondly, PRP specifically).

To some extent – but again, not exactly – this sequence mirrors the common social science distinction between structure and action. The contextual features identified higher up in the table tend to be predominantly structural – the politico-administrative culture and the structure of the political system. Those towards the end of the table are more concerned with contemporary actors and actions, such as resistance from stakeholders or the presence of international pressure to reform. Of course this distinction between structure and action is not absolute (as many social theorists have affirmed). The two are intimately connected in various

ways. Structures commonly set the conditions for actions, as when a hospital manager finds s/he cannot persuade the medical staff to accept a proposed new system for auditing the efficiency of clinical care because they see it as contradicting the culture of clinical autonomy. Equally, dynamic actions (such as reformers seizing ‘windows of opportunity for reforms’) may set in train a longer term shift in cultures or organizational structures. Leadership is a good example of a factor which can appear as either a structural or an action element in explanations. If a culture or political system does not endorse the idea of senior public servants as leaders – seeing them as mere cogs in the machine or just as politicians’ day-to-day ‘fixers’ – then the system will lack leadership capacity – a structural weakness. In other analyses, however, a particular public service leader or group of leaders may be seen as having ‘made the difference’ by giving a reform legitimacy and continuous support over time. Here leadership is treated as an action factor.

Most, but not all, the identified contextual influences can work both ways, in the sense that they may facilitate a particular reform, or inhibit it. A centralized strong executive in a majoritarian system may enable one country to reform rapidly and on a wide scale – *if* the executive in power supports that kind of reform. If it does *not* support that type of reform then exactly the same structural feature – a strong central executive – may act as a barrier to that type of reform. Taken as a whole, the literature probably treats contextual influences in a somewhat biased way. They seem to get more attention when they act as barriers, or pitfalls for reform. When they facilitate reform they may not even be mentioned – it is simply assumed that the reform was a good idea and that, when implemented, it worked. This hypothesized bias has the effect of understating the overall influence of contextual factors, because it underplays them when they are positive.

A distinction needs to be made between contextual influences affecting the implementation of reforms versus contextual influences on the impacts of (implemented) reform. Few studies in our database attempt to differentiate between the two, in part because it is difficult methodologically and in part because it is ambiguous conceptually. However, simply because a reform is *successfully* implemented does not necessarily imply that this same reform is producing positive results. As noted in Table 10, successfully implemented reforms have in some cases been found to lead to unintended, negative results. Therefore one needs to clearly define what constitutes success – in this case the extent to which a reform is implemented *or* the extent to which it produces positive results, or both.

Many of the studies in the database mention contextual features as affecting the *implementation* of reform. However, most of them fail to focus on or explain the precise

mechanism by which the identified contextual factors affect the *impacts* of implemented reform. For instance, the studies mentioning sudden accidents or scandals as important influences or the studies which found opposition from labour unions to be a key factor, typically sought to explain implementation failure, not impacts *per se*. The same appears to be the case for factors such as the structure of the political system, rapid turnover of governments, cultural resistance to change and external pressure for reform. Other types of factors, however, more closely reflect the impacts of reform. These include the socio-demographic characteristics of users and the standardization of service provision, which had a clear impact on the dissatisfaction of service users and had little, if anything, to do with implementation. A third category exists which more closely reflects both implementation and impacts, including factors such as politico-administrative culture, lack of current administrative capacity, openness to and experience with the private sector (which may affect both implementation of reform and the way in which it is carried out, for instance, by how customer focused service provision is), the role of leadership and management or the degree of acceptance of reform measures.

A close reading of the studies in the database suggest that we can already make a number of context-related, middle range generalizations about certain typical elements within NPM reforms. For example:

- Performance related pay requires a long list of supportive local conditions before it stands a good chance of working as intended. In particular it tends not to work well where a) political patronage determines most senior appointments, b) the bonuses available are only a very small % of the total remuneration, c) the work is hard to measure in an objective and widely-accepted way, d) there is a lack of managerial capacity in the ministry promoting it and/or in the ministries/agencies where it is implemented, e) real or perceived issues of favoritism and corruption exist (and so on) (e.g., Jeannot and Guillemot, 2010; Makinson, 2000; Marsden and French, 1998; Randma-Liiv, 2005; World Bank, 2001).
- Performance targets can definitely have a significant impact on performance, but usually only where backed up by significant penalties/incentives (Andersen and Pallesen, 2008; Kelman and Friedman, 2009; Pollitt et al., 2010; Propper et al., 2008). They are likely to lead to negative results and prove unsustainable in the medium and long run if they are improperly defined and understood and if they are implemented in a highly politicized context (Nemec, 2007). In such cases it is not necessarily that performance targets actually disappear, but rather that they are in a state of almost constant 'churn', changing substantially from year to year and failing to provide a stable basis for year-on-year planning and improvement.

- Contracting out often fails to yield significant savings, particularly in the medium and long term. Again, the context has to be right for contracting out to work in the way it is supposed to. Necessary supports include a) the possession of contract design and management skills, including strategic planning skills, by the staff of the parent public sector organization, b) the presence of real competition (which may exist at first but then disappear because of private sector mergers and take-overs) and c) that the activity being contracted is one which can be specified in fairly firm detail (Atun et al., 2005; Macinati, 2006; National Audit Office, 2009; Nemeč and Kolisnichenko, 2006).
- Use of market-type mechanisms (MTMs) may work better in simpler, more standardized services than in complex, unstandardized, professionalised human services. There are a number of studies in a number of countries indicating problems with the application of MTMs to healthcare and educational services, such as inadequate financial resources and policy and managerial capacity (particularly in countries in CEE, but not exclusively in CEE) (Atun et al., 2005; Nemeč, 2008).
- The politico-administrative culture is mentioned as a significant, often overarching, factor shaping reform impacts in many studies – especially in Central and Eastern Europe, but also in the West (Brunetta, 2009; Capano, 2003; Læg Reid et al., 2008; Nemeč, 2007; OECD, 2010; Societatea Academică Română, 2010; Şandor and Tripon, 2008; Torres and Pina, 2004).
- The structure of political system may be facilitating or inhibiting for rapid or large-scale reform. The incremental decision-making process in, say, Norway or the Netherlands is more likely to inhibit such reform whereas the centralized, majoritarian system in the UK and a more consensual one is more likely to facilitate reform (Askim et al., 2010; Christensen, 2001; Minvielle, 2006; Norges Forskningsråd, 2007; Pollitt and Bouckaert, 2011; Pollitt et al., 2010; Roberts, 1997).
- Strong leadership can be a critical factor supporting the introduction and implementation of NPM-type reform measures, but the measures themselves, although fuelled and supported by strong leadership, may nevertheless produce mixed results, depending on other factors (Atun et al. 2005; Edelenbos and Klijn, 2009; Hammerschmid et al., 2011; Laforgue, 2004; Promberger, 2008; Verhoest, 2005; Yesilkagit and de Vries, 2002). In a number of studies a supporting factor has been found to be the internal and external legitimacy or acceptance of the proposed measures (Hondeghem and Depré, 2005; McNulty and Ferlie, 2004; Nagyistók, 2010; Pollitt et al., 1998b; World Bank, 2001).

## **12. An illustrative example**

It may be worth illustrating the challenges of evaluating the impacts of NPM-type measures by looking in more detail at one particular example. For this purpose we have deliberately chosen a very strong example – one of the few studies that include a rigorous analysis of a large statistical database that covers mostly changes in process and output, but also, to a limited extent, in outcomes. It is probably no coincidence that this comes from the healthcare sector (where measurement is professionally endemic) and from the UK (where NPM reforms have been both frequent and intense).

Kelman and Friedman (2009) focused on the attempt within the UK National Health Service to reduce waiting times in hospital accident and emergency (A&E) departments. They were especially concerned to see whether the apparent improvements brought about by a vigorous performance targeting regime had also led to dysfunctional ‘side’ effects. The specific target they investigated (one among many – see Pollitt et al., 2010) was that in A&E departments 98% of all patients would be treated within four hours of arrival. This therefore concerned a process (service users passing through Accident and Emergency departments) and a set of outputs (were those users treated by the medical and nursing staff?). Surveying 155 hospital trusts Kelman and Friedman found that the percentage achieving this target was 1.24% in the third quarter of 2002 but had risen to 59.4% by the third quarter of 2007. This very marked improvement was attributed to the government campaign, commencing at the beginning of 2003, which featured the target as part of an overall system for publicly rating hospitals as having three, two, one or zero ‘stars’, and which later offered hospitals cash incentives for meeting the target. No evidence of any dysfunctional effects was found.

This was a particularly sophisticated, detailed and careful study. It demonstrated, beyond reasonable doubt that waiting times were dramatically reduced, and that certain kinds of possible dysfunctions did not appear to take place. For example, there was no evidence that quality of care decreased (p.17), indeed, the Accident and Emergency-related death rate slightly decreased. Nor did the increased efforts in A&E seem to detract from performance in related departments – there was no corresponding increase in wait times in orthopaedics, for example (p. 17). Nor was the improvement a once-only ‘blip’, focused on the month in which the government collected its performance statistics. The measurement period did produce a short-lived ‘spike’ in performance, but afterwards it settled down at a much higher level than before (p.18). In short, the regime of tough performance targets, intensively monitored by central government, with the results openly published, appeared to work.

Yet even here, there is room for alternative interpretations. To begin with – as Kelman and Friedman themselves acknowledge – ‘both budgets and staffing for the NHS, including A&E departments, increased significantly during this period...and we do not claim that all the overall improvement reflected in performance was due to attention to the target’ (p. 929). So perhaps this was not wholly or even mainly an efficiency improvement, because better outputs were partly or wholly explained by more inputs?

More seriously though, close to the time that the Kelman and Friedman article was published, so was the report of an inquiry into events at Mid Staffordshire NHS Foundation Trust (Healthcare Commission, 2009). This trust hospital had come into the news because of extraordinarily high mortality rates, and evidence of widespread dissatisfaction among patients. The Healthcare Commission concluded, *inter alia*, that ‘there were deficiencies at virtually every stage of the pathway of emergency care’, and that ‘Doctors were moved from treating seriously ill patients to deal with those with more minor ailments in order to avoid breaching the four-hour target’ (Healthcare Commission, 2009, p. 129). This obsession with breaching the target was not the only management failing – the inquiry found a number of others – but it was certainly one of them. Over a three year period Mid-Staffordshire had suffered between 400 and 1200 more patient deaths than would have normally been expected. So we seem to have two studies, each thorough, pointing in different directions. On the one hand, an academic statistical analysis of 155 hospital trusts finds ‘no evidence for any of the dysfunctional effects that have been hypothesized in connection with this target’ (Kelman and Friedman, 2009, p.917), while on the other hand, a detailed official review, based on extended investigation ‘on the ground’ (300 interviews and 1000 documents) finds plenty of evidence of just such distortions and dysfunctions. Of course, one could say that Mid-Staffordshire was only one hospital trust out of 155, and was quite exceptional. One could also point out that the big improvement in national average waiting times in A&E had presumably saved an unquantified number of lives. On the other hand, 400-1200 excess deaths is rather a large exception and, at the very least, points to the limitations of a purely statistical approach to organizational analysis. A later detailed inquiry into Mid Staffordshire concluded that ‘there can no longer be any excuse for denying the enormity of what has occurred’, and that ‘a high priority was placed on the achievement of targets, and in particular the A&E waiting time target. The pressure to meet this generated a fear, whether justified or not, that failure to meet targets could lead to the sack’ (Francis, 2010, pp. 3 and 16). Mid Staffordshire seems to have been an unusually tragic example of – in our terms – true effectiveness being abandoned in the pursuit of a particular kind of measured efficiency. It is also a case where particular, local contextual factors seem to have led to a perverted management response to performance

targets, whereas exactly the same targets in other hospitals helped to stimulate genuine performance improvements.

Overall, therefore, one may incline towards a mixed assessment. Kelman and Friedman showed that, across the country, the ‘blitz’ on A&E waiting times produced major gains with few apparent clinical penalties or other distortions. The tragic events at Mid-Staffordshire show that, in one particular place, this improvement mechanism could go horribly wrong, and that even a sophisticated statistical analysis at the national level would not necessarily pick that up. More widely, however, one may take note of the huge effort that was required to achieve this particular change in just one (admittedly very important) aspect of hospital operations. It required a massive and detailed national database, careful monitoring and checking, and very strong political pressure, including financial incentives and threats to top managers’ jobs in the case of failures. Not every jurisdiction has the skills and resources to mount such a policy.

### **13. Implications for further research**

The above discussion and findings hold out some clear implications for further research on the impacts of public management reforms. These can be summarised as follows:

1. Researching the full impact of a major reform is frequently difficult and challenging. It will usually demand the commitment of researchers and resources over a considerable period of time. Broad, impressionistic overviews have their place – not least that they may stimulate more detailed research – but ‘getting to the bottom of things’ takes time, access, analytic skill and adequate resources. It is usually hard for an individual researcher to carry out this kind of work – the capacities of a dedicated, multi-skilled team (such as COCOPS) are far more likely to achieve firmly warranted insights.
2. ‘Getting to the bottom of things’ includes seeking to understand both what and how contextual factors affect implementation of reform, on the one hand, and outputs and outcomes of reform, on the other. A reform measure or reform package cannot necessarily be deemed successful solely on the basis of how successful implementation is – it is also necessary to take into account the outputs and outcomes which that reform produces.
3. Identifying impacts is much less difficult in countries and sectors where there is already the habit and practice of collecting routine performance data (see, e.g. National Audit Office, 2001). Researchers can then use that and build on it. They do not face the heroic – often impossible – task of assembling systemic ‘before and after’ performance data themselves. Such habits and practices have hitherto been far

more common in certain sectors and countries than others. The generation of high quality, publicly-available performance information has become routine in the highly professionalized sectors of healthcare and education, and is more widely visible in the UK and the Nordic group of countries than in, say, Belgium, Germany or Spain, or in most countries in CEE, where Estonia for instance is an exception (see Pollitt and Bouckaert, 2009, for an Anglo-Belgian comparison). However, even where plentiful routine data is available there is often still plenty of hard work for the researcher to do, especially if the objective is to estimate the effects of a range of reforms over a period of time (e.g. Carrera et al., 2009)

4. The independence of the researcher may well be a significant influence on the strength and relevance of the findings. Although we have studied many official reports as well as academic studies, most (not all) of the studies which have been willing and able to draw robust conclusions about outcomes and outputs have been academic.
5. The priority for the future should be the direct study of changes in outputs and, especially, outcomes. For the public these are the real ‘results’ of reform, but they have been far less often studied than changes in organizational structures and processes (that, of course, is partly because they are more difficult to study, both practically and, sometimes, politically). That this is an old message (it was already being delivered in the 1980s) is one sign that it is difficult to do. However, our database has a modest number of studies that show that sometimes, at least, it is achievable. The findings from such studies are particularly valuable
6. There is no ‘one best way’ of studying outputs and outcomes. Tight statistical studies (e.g. Kelman and Friedman 2009; Wilson and Piebalga, 2008) and the more broad-scope synthetic analyses with qualitative elements (e.g. Baggot, 1997; Kuhlmann et al., 2008; Nemeč and Kolinichenko, 2006) can each yield strong and valuable conclusions. Where similar countries or organizations can be compared it can sometimes be possible to conduct telling ‘natural experiments’ (e.g. Propper et al., 2008). For a variety of reasons it may be both more feasible and more desirable to undertake experimental approaches to public management research than in the past (James, 2011; Margetts, 2011).
7. The basic definition of performance concepts such as ‘efficiency’, ‘effectiveness’ and ‘outcomes’ always needs to be done with great care, and then rigorously adhered to.

#### **14. Concluding observations**

What, finally, can we say about the impacts of NPM in Europe? The situation appears paradoxical. On the one hand there have been endless publications – both academic and

official – concerned with NPM-like programmes and techniques. Yet, on the other, our solid, scientific knowledge of the general *outcomes* of all this thinking and activity is very limited. Claims and counterclaims outnumber hard, carefully collected evidence – and by a substantial margin. That was the case in the mid 1990s (Pollitt, 1995) and it remains the case today. One of the scholars who originally ‘discovered’ and defined NPM, when writing a recent review of a large new edited volume on the subject, put it like this:

‘Indeed, what will surprise many readers is how little we seem to know after decades of research about whether and how far NPM “worked” in what is commonly said to have been its main original concern, namely to cut costs and improve efficiency’ (Hood, 2011, p. 738)

A summarising metaphor might be that there is an ocean of studies of the application of NPM ideas within the Europe, but only a modest sea of works that offer direct empirical analysis of *outputs*, and no more than a small pond that convincingly connect specific reforms to particular *outcomes*.

It is true that our database holds a good deal of interesting information about changes in *outputs* resulting from NPM, but what the collectivity of this knowledge suggests is that these vary enormously, from highly positive through ‘no change’ to rather negative, depending on circumstances. And the circumstances (contexts) seem to be very various. That is why exploration of these contextual influences has been an important focus for us. The majority of our publications, however, get no further than discussions and analyses of changes in *processes and activities*. That is the ‘ocean’. To put it in a challenging way, we would, on this basis, argue that all generalisations of the form ‘NPM reforms of type X lead to outcome Y’ are suspect, at least in the sense that the mountain of evidence in our database will not yield any such firm and overarching conclusion. We may note, *en passant*, that multiple difficulties – both conceptual and empirical – in evaluating the impacts of public management reform are neither new nor confined to the Europe. More than a decade ago a leading authority on the famous NPM reforms in New Zealand noted that:

‘All too frequently important methodological issues are ignored or glossed over. As a result, claims are made about the impact of particular management changes for which the evidence is either thin or of questionable validity’ (Boston, 2000, p. 25)

and, furthermore:

‘there are serious difficulties in securing the necessary evidence to assess many aspects of the New Zealand reforms’ (*ibid.*, p. 38)

If the picture we arrive at after a meta-analysis of hundreds of documents ostensibly concerned with the impacts of NPM in Europe is similarly patchy and problematic, what

conclusions can be drawn? Broadly speaking, there are two obvious possibilities. First, perhaps there *are* general rules which adequately summarize the impact of NPM reforms, but we have not found them. Alternatively, second, there may be no such general rules, in which case we may have been asking the wrong kind of question.

Let us consider the first alternative first. Why might our research have failed to find the systematic general connections between NPM reforms (reform 'inputs' if you will) and the outputs and outcomes of activities conducted by public organizations? Here there are at least two sub-alternatives. One is that our literature search was incomplete, and that there *are* documents out there, somewhere, that would yield more definite general connections and rules. This is a possibility. The literature is huge and very diverse, and it is always possible that we and our partners and collaborators have overlooked some crucial analysis. If so, we hope that those who know of such missing links will let us know. On the whole, though, we find the probability of the existence of some missing, yet definitive analysis - something that will provide a *general* rule or formula – fairly low. (That we have missed *something* is almost certain, but that we have missed something *crucial* is unlikely.) Further, even if some 'killer study' does exist, it is evidently unknown by the hundreds of authors of the documents in our database – so we would not be the only ones 'in the dark'.

Another possibility is that useful knowledge exists, but that it is hard – or even impossible – to put into a codified, explicit, 'scientific' form. It may be a form of 'craft' or tacit knowledge, which experienced practitioners have developed but which depends on un- or seldom-articulated understandings of the nuances of particular situations and formations of reform actors. In fact some writers have suggested that this type of knowledge is of great (but largely unacknowledged) importance in organizational practice (Tsoukas, 2005). If, however, this is the case then, *ipso facto*, we are not going to find the general rules by searching academic and official literature. One would have to adopt quite different research strategies, for example by using prolonged participant observation of experienced public managers at work.

The second main possibility is that there are no general rules to be had. It is vital to understand that this does *not* mean that we can have no knowledge of NPM reform impacts. It simply means that the knowledge is unlikely to come in the form of a general, 'Z follows Y follows X' rule. The topic is just too complex and contingent on many varying factors and influences for that kind of stable general relationship to be available. However, more specific, context-dependent or *ad hoc* forms of knowledge may well exist. And that, we would argue, is one of the things that our meta-analysis points towards. In this respect the findings about

contexts are crucial. It is by carefully sifting those that more particular and conditional (and local) sets of relationships may be identified. Thus limited generalizations – bounded by time and place, and conditional upon the presence or absence of certain contextual factors – would still be entirely feasible. And if they *are* possible (as we believe they are) they are also likely to be extremely useful to practising reformers.

Our analysis thus far has certainly confirmed the importance of contextual factors (section 11 above). We have suggested a framework for analysing these factors and offered a wide range of examples, drawn from the database, of contextual influences in action. Yet we are conscious that this still amounts to only a few preliminary steps in the process of illuminating the full complexities of context. Even so, it seems appropriate in these concluding observations to identify some of the implications of what we have learned about context thus far.

A first, rather practical and basic implication would seem to be that reformers would be well-advised always to include in their reform teams members with up-to-date local knowledge, both formal and tacit. Such individuals should be best-placed to anticipate problems with the local culture, standard operating procedures, staff morale, tendencies to corruption, and so on. Reform teams composed entirely or primarily of consultants or reform enthusiasts from outside the target organizations or programmes are inherently risky.

A second implication is that some kind of *ex ante* systematic review of likely contextual factors constitutes a highly desirable feature of any reform plan. In this context our database can provide, at the minimum, a kind of check-list of possible items to think about. These would include, for example:

- Considering the implications of the wider political system. The overall political system (broad scope) does not usually change much in the short term: reforms may therefore need adapting according to whether they are being inserted to, say, an aggressively competitive, majoritarian system or a more consensually-oriented multi-party system)
- Assessing the likelihood of sudden turnovers of governments and/or restructurings of governmental structures and functions. Such volatilities may exercise a major influence on the continuity and cohesion of reform efforts in time and across governmental levels and units. This factor is more difficult to predict as unanticipated accidents, crises or scandals may result in such *voltes faces*. However, in the case of some countries (Italy in the 1980s and '90s, some CEE countries since the mid 1990s) rapid changes of government are a clear and expected pattern.

- Assessing the local organizational culture. This is a longer term influence, hard to change radically in the short term but potentially malleable in the medium to long term
- Recognising the degree of corruption existing in the particular sector/organization. Many NPM reforms are hard to implement in a high corruption environment because they involve giving managers and front-line staff greater autonomy – in cases where corruption is prevalent it may be necessary to tackle corruption, at least in selected areas or sectors, before NPM reforms can be fully implemented. Corruption is one particular aspect of culture, and is similarly hard to transform in the short term, although it can be progressively reduced over time
- Assessing the existing capacity and skill set of the organization(s) which are to undergo reform. Skill sets can be altered in the medium term if a planned programme of training and recruitment is put in place. In the short term there is little point introducing a new technique (e.g. regular appraisal interviews; accruals accounting) if enough staff are not trained to operate them
- Analyzing the socio-demographic situation of users and citizens that are targeted by reform, since different groups may experience the effects of reform differently and, conversely, the impacts of reform may be different, based on different socio-demographic characteristics
- Considering the role and influence of more 'organized' factors affecting implementation and possibly impacts per se, such as labor unions or (senior) civil servants who are opposed to reform and can possibly exert significant influence.

A third implication is clearly that those researching public management reform need to be particularly careful to specify the domain over which they consider their findings are likely to hold reasonably true, and to identify the main contextual influences they see as being in play within that domain. Too many of the documents in our database are loose or even silent about these issues of domain and context. Only by being more specific about these issues (and, indeed, by actively *theorizing* them) can it be hoped that sound, contextually grounded, middle level generalizations will emerge.

There remains at least one puzzling and fascinating question which can never be answered by the study of our database alone. It is the question of why such a huge amount of reform - organizational change and upheaval in almost every European state – has taken place *if* the evidence for its positive effects on citizens is so slender? On the basis of the works we have

reviewed we can attempt an informed speculation, but no more. There are at least four candidate explanations:

1. NPM has been primarily an issue of faith rather than demonstrated results (Pollitt, 1995). Politicians and senior civil servants have somehow been 'sold' a set of ideas and principles which are actually by no means as widely efficacious as their proponents have claimed. At least one leading scholar has likened contemporary public management doctrines to a religious belief (Hood, 2005). And many writers have used the popular term 'guru' when describing certain leading management thinkers (e.g. Jackson, 2001). More analytically, perhaps, NPM has been seen as one manifestation of a broader ideology of managerialism (Pollitt, 1990) or as a collection of doctrines (Hood and Jackson, 1991). In each case NPM is conceptualised as something that goes far beyond specific instruments, standard techniques and actual, observable performance. Rather it is seen as part of a wider pattern of values, ideas and beliefs that are carried and spread by certain groups (Hartley, 1983; Sahlin-Andersson and Engwall, 2002).
2. NPM has been adopted primarily for its symbolic properties. It symbolizes modernization; it symbolizes a more customer-oriented stance by the public authorities and it also symbolizes a populist, anti-bureaucratic stance. In the short term it can be used by political leaders as evidence of an active, reforming government (Box et al., 2001; Common, 1998). It can also endow modernizers with status and influence within the bureaucratic machine, whether or not the eventual outputs and outcomes are particularly positive (Sundström, 2006). There is thus an important sense in which symbols are not merely symbolic – on the contrary, being able to identify oneself with the 'right' symbols can confer status and influence.
3. NPM has been a somewhat self-interested and even, on occasions, slightly cynical exercise by politicians and senior civil servants. Faced with high pressures to restrain ever-rising public expenditure, governments have launched schemes to cut back the public sector but, knowing these are unlikely to be popular, they have cloaked them in programmes of reform that promise 'more with less'. Thus the main aim has actually been economy and/or increased control by the top over operating agencies, not improved customer service (though the latter is quite an acceptable secondary effect on those occasions when it can be achieved). For example, looking at citizens' charters in the UK, Falconer and Ross suggest that 'charters, in practice, do serve primarily as, and are best viewed as, managerial instruments and not as mechanisms for the empowerment of service users' (1999, p. 350). More recently, we have seen a UK government issue a white paper on management reform which claims to be going to 'wrest power out of the hands of highly paid officials and give it back to the

people' whilst scarcely mentioning the fact that the reform programme is actually driven by the need to make historically unprecedented cuts in public spending (Minister for Government Policy, 2011, p. 5). A major part of the substance of the proposed reforms seems likely to be a further round of NPM-style contracting out.

4. NPM has been pressed onto some countries by other countries or institutions – particularly by the original, core NPM enthusiasts, namely Australia, New Zealand, the UK, the USA and international organizations, especially the OECD, SIGMA and the World Bank (see, e.g. Arndt, 2008, Independent Evaluation Group, 2008; OECD 1995, 2010; Nunberg, 1999; Pollitt, 2010). The World Bank and the OECD, in particular, have at various times propounded models of 'good governance' or of the preferred direction of public sector reform. So NPM type reforms have not always been freely 'chosen'. Sometimes they have been requested or strongly recommended or even, for some states, especially in the CEE and the developing world, 'required'. The intensification of the Eurozone fiscal crisis during 2011 seems to have led to further examples of this. For instance, in November EU Commissioner Oli Rehn presented the Italian government with a letter detailing a long list of points where the EU expected reform action, many of which were either directly administrative or had significant administrative implications.

It is important to notice that these four explanations are not at all mutually exclusive. They can be additive in one country, or present in quite different mixtures in different countries. For example, in the CEE countries explanations 2 and 4 often appear to fit quite well. Or in the recent conditions of fiscal crisis one can perhaps detect elements of 3 in countries such as France, Italy or the UK (e.g. Walker, 2011) or 4 in countries such as Ireland, Italy or Greece. Or again, explanation 1 (faith) seems to fit quite well for certain leaders such as Margaret Thatcher (UK Prime Minister 1979-89) or Ronald Reagan (US President 1980-88).

As indicated above, each explanation has found some supporters in the literature. But each immediately prompts further sub questions of its own. In the first case, who has sold NPM so effectively? Who have been the principal 'carriers' of these doctrines? Some writers have traced this back to neo-liberal theorists such as Niskanen (1973) or Downs (1967), but it seems unlikely that academic ideas by themselves could have spread so far and fast without more powerful support. Management consultancies are one possibility (Sahlin-Andersson and Engwall, 2002; Saint-Martin, 2000) but it seems obvious that we must also count politicians themselves as both enthusiastic 'buyers' and then, subsequently, 'sellers'. The core NPM idea that the public sector is deeply inefficient and that therefore, with the right business techniques, money can be saved *and* services can simultaneously be improved is obviously a deeply attractive one to many politicians. And most of them may not be in a very strong position to assess its underlying accuracy: with the rise, in many EU states, of the

'professional politician' (men and women who have never done much of anything else) the amount of experience of running large, complex organizations which exists within the political elite – never large – has probably fallen even further (Mottram, 2008). In any case, even in the private sector, the ability to convince oneself that *this time* a reform really *will* work has for long been a widespread and necessary ingredient of managerial life (Brunsson, 2002).

For the second explanation (NPM as an exercise in symbolism) the question arises of how and when the dominant set of symbols change. It appears, for example, that, although NPM became dominant in the 1980s and '90s, the new paradigms of 'governance', 'networks' and 'partnerships' have emerged to rival or even supersede NPM in the period since the mid 1990s (Koppenjan and Klijn, 2004; Osborne, 2010, Pollitt and Bouckaert, 2011, Pollitt and Hupe, 2011).

For the third explanation (manipulation by politicians and senior officials whose prime motive is actually to save money and/or gain closer control over operating agencies) one question is 'for how long can this kind of public relations exercise continue to be effective?'. The answer is likely to vary from country to country. One might expect that in countries which had already experienced long histories of repeated reforms (such as France, Sweden or the UK – or some of the CEE states) there would be a lower willingness (among both citizens and public sector 'insiders') to believe that 'this time' there really would be a transformation of the quality of service experienced by the citizen (see, e.g. Pollitt, 2007). Certainly there is some evidence that the public in these countries are increasingly skeptical of government figures claiming performance has improved (e.g., Fellegi, 2004; Holt, 2008; Magee et al., 2003). It is also the case that in a number of countries the mass media provide a distinctly hostile environment for those who wish to celebrate successful reform:

'Mass media that are characterised by a combination of politico-economic antagonism towards public services and journalistic cynicism about politics form a difficult setting for the publication and celebration of "success stories"' (Clarke, 2005, p. 226)

The fourth explanation (pressures from outside) invites more detailed questions about how this pressure is powered and organized? What levers and incentives are in operation, persuading or obliging reluctant or agnostic countries to adopt NPM-type reforms? Clearly in the CEE countries there were at least two processes continuing side-by-side. First, there was gaining money and assistance from international agencies such as the World Bank or the European Commission. Second, there was the anxiety, prior to EU enlargement, to meet the

requirements of the club of EU member states, so as to be allowed in. It may be significant that, once inside the club, a number of CEE countries appear to have slackened off on their public management reform programmes (though these include non-NPM as well as NPM measures – Meyer-Sahling, 2009, p.7; Verheijen, 2007, p. x).

Taken together, then, there are a range of mutually re-inforcing reasons why the state of affairs documented in this meta-analysis is able to continue to exist. We cannot here go further in disentangling or confirming the parallel explanations – more research is needed, and it will need to be of a different, more probing type than our own database. We can, however, note that all these reasons are already in academic currency and that, separately and collectively, they possess a certain face-validity. Management *reform*, it seems, is far from being simply applied management *science*.

We conclude with an observation from a 2008 address by a retiring British Permanent Secretary. Reflecting on 15 years at the top, including a period as the Permanent Secretary responsible for public management reform, he said:

It is debatable how far looking at issues through the prism of outcomes and outputs has taken hold in terms of either policy making or of political rhetoric – it is just so much easier to boast of more money or more tangible extra things, whether teachers, police on the beat, or new schools, rather than to describe outcomes’ (Mottram, 2008, p. 2)

Whilst we fully agree with Sir Richard Mottram that it is much easier to trade in other evidence rather than outcomes we also hope that one of the ‘value-added’ functions of academic research such as the COCOPS project is that a focus on outcomes and outputs, however difficult, can be sustained. That is the spirit in which we offer readers this report.

## Bibliography

- Andersen, L.B. & Pallesen, T. (2008) Not just for the money? How financial incentives affect the number of publications at Danish research institutions, *International Public Management Journal*, 11(1), 28-47.
- Andrews, R., Boyne, G. & Walker R. (2011) The impact of management on administrative and survey measures of organizational performance, *Public Management Review* 13(2), 227-255.
- Arndt, C. (2008) The politics of governance ratings, *International Public Management Journal* 11(3), 1-23.
- Askim, J., Christensen, T., Fimreite, A. L., & Læg Reid, P. (2010) How to assess administrative reform? Investigating the adoption and preliminary impacts of the Norwegian welfare administration reform, *Public Administration*, 88(1), 232-246.
- Atun, R., Menabde, N., Saluvere, K., Jesse, M. & Habicht, J. (2005) Introducing a complex health innovation – primary health care reforms in Estonia (multimethods evaluation), *Health Policy* 79, 79-91
- Audit Commission (2009) *Final score: the impact of Comprehensive Performance Assessment of local government, 2002-2008*, London: The Audit Commission.
- Baggot, R. (1997) Evaluating health care reform: the case of the NHS internal market, *Public Administration* 75(2), 283-306.
- Barzelay, M. (2000) The New Public Management: a bibliographical essay for Latin American (and other) scholars, *International Public Management Journal* 3, 229-265.
- Bauer, M. (2009) Impact of administrative reform of the European Commission: results from a survey of heads of unit in policy-making directorates, *International Review of Administrative Sciences* 75(3), 459-472.
- Belorgey N. (2010) *L'hôpital sous pression, enquête sur le nouveau management public*, éd. La Découverte, Paris
- Bodnár, V., Drótos, G. Kiss, N. & Révész É. (2005) Performance measurement in a transforming university setting. EIASM Conference: 3rd Conference on Performance Measurement and Management Control, Improving Organizations and Society, 22-23 Sept., Nizza.
- Boston, J. (2000) The challenge of evaluating systemic change: the case of public management reform, *International Public Management Journal* 3, 23-46.
- Boston, J. et al (1996) *Public management: the New Zealand model*, Auckland, Oxford University Press.
- Box, R., Marshall, G., Reed, B. & Reed, C. (2001) New Public Management and substantive democracy, *Public Administration Review* 61(5), 608-619.
- Boyne, G., Farrell, C., Powell, M. & Walker, R. (2003) *Evaluating public management reforms*. Buckingham: Open University Press.
- Brunetta, R. (2009) Reforming the public administration to make Italy grow, *Review of Economic Conditions in Italy* 3, 337-368.
- Brunsson, N. (2002) *The organization of hypocrisy: talk, decisions and actions and organizations*, (2<sup>nd</sup> edition), Copenhagen, Abstrakt Liber, Copenhagen Business School Press.
- Capano, G. (2003) Administrative traditions and policy change: when policy paradigms matter. The Case of Italian administrative reform during the 1990s, *Public Administration* 81, 781-801.
- Carrera, L., Dunleavy, P. & Bastow, S. (2009) *Understanding productivity trends in UK tax collection*, London, LSE Public Policy Group Working Paper, July.
- CEEP (2010) *Public services within the European Union and the 27 member states: statistics, organizations and regulations*. Paris: project with support of the European Commission, May.
- Centraal Planbureau (2007) Effectiviteit van de politie: van meer naar beter blauw, in MEV 2007 Speciale Onderwerpen, Centraal Planbureau, 139-150.
- Chan, J., Ho-Pong, T. & Chan, E. (2006) Reconsidering social cohesion: developing a definition and analytical framework for empirical research. *Social Indicators Research*, 75(2), 273-302.

- Christensen, T. (2001) Administrative reform: changing leadership roles, *Governance* 14/4, 457-480.
- Christensen, T. & Lægreid, P. (eds.) (2001) *New public management: the transformation of ideas and practice*, Aldershot, Ashgate.
- Christensen, T., Lægreid, P., & Stiger, I.M. (2006) Performance management and public sector reform: the Norwegian hospital reform, *International Public Management Journal* 9(2), 113-139.
- Clarke, J. (2005) 'Performing for the public: doubt, desire and the evaluation of public services', pp. 211-232 in P. Du gay (ed.) *The values of bureaucracy*, Oxford, Oxford University Press.
- Clifton, J. & Díaz-Fuentes, D (2010) Evaluating EU policies on public services: a citizens' perspective. *Annals of Public and Cooperative Economics*, 81(2), 281-311.
- Common, R. (1998) Convergence and transfer: a review of the globalization of the New Public Management *International Journal of Public Sector Management* 11:6/7, 440-450.
- Cooper, H. (1998) *Synthesizing research*. Thousand Oaks: Sage.
- Dankó, D. & Molnár, M. P. (2009) Stratégiai szemlélet a gyógyszer-támogatásban I. rész Célok, eredmények, eszközök, IME VIII. Évfolyam 10. Szám, December.
- Divay, S. (2009) Nouveaux opérateurs privés du service public de l'emploi: les pratiques des conseillers sont-elles novatrices?, *Travail et Emploi*, (3)119, 37-49, juillet-septembre.
- Dorrell, S. (1993) *Public sector change is a world-wide movement*. Speech by the Financial Secretary to the Treasury, Stephen Dorrell, to the Chartered Institute of Public Finance and Accountancy, London, 23 September.
- Downs, A. (1967) *Inside bureaucracy*, Boston, Little Brown.
- Dunleavy, P., Margetts, H., Bastow, S. & Tinkler, J. (2006a) Digital-era governance: IT corporations, the state and e-government. Oxford: Oxford University Press.
- Dunleavy, P., Margetts, H., Bastow, S. & Tinkler, J. (2006b) New Public Management is dead – long live digital-era governance. *Journal of Public Administration Research and Theory* 16(3), 467-494.
- Edelenbos, J. & Klijn, E-H. (2009) Project versus process management in public-private partnership: relation between management style and outcomes, *International Public Management Journal* 2(3), 310-331.
- Emery, Y. (2004) Rewarding civil service performance through team bonuses: findings, analysis and recommendations, *International Review of Administrative Sciences*, 70(1), 157-168.
- Emery, Y. & Giaouque, D. (2001) New Public management, contracts for the provision of services and financial incentive mechanisms: practices and problems encountered in Switzerland, *International Review of Administrative Sciences* 67, 151-171.
- Falconer, P. & Ross, K. (1999) Citizen's charters and public service provision: lessons from the UK experience, *International Review of Administrative Sciences* 65, 339-351.
- Federal Ministry of the Interior (2006) *Focused on the future: innovations for administration*. Berlin: Federal Ministry of the Interior.
- Fellegi, I. (2004) Official statistics – pressures and challenges, *International Statistical Review* 72(1), 139-155.
- Ferlie, E. & Geraghty, J. (2005). Professionals in public service organizations: implications for public sector reforming. In E. Ferlie, L. Lynn Jnr. & C. Pollitt (Eds), *The Oxford handbook of public management*, pp. 422-445. Oxford: Oxford University Press.
- Flyvberg, B. (1998) *Rationality and power: democracy in practice*, Chicago, University of Chicago Press.
- Francis, R. (2010) *Independent inquiry into care provided by Mid Staffordshire NHS Foundation Trust, January 2005-March 2009*, HC375-1, London, The Stationary Office.
- Gore, A. (1997) *Businesslike government: lessons learned from America's best companies*. Washington, DC: National Performance Review.

- Gow, J. & Dufour, C. (2000) Is the New Public Management a paradigm? Does it matter? *International Review of Administrative Sciences* 66(4), 573-593.
- Green, A. & Janmaat, J. G. (2011) *Regimes of social cohesion*, London, Palgrave/Macmillan.
- Hammerschmid, G., Krischok, A. & Steigenberger, K. (2011) Agencification experiences in Austria. In Bouckaert, G., Laegreid, P., Verhoest, K., Rommel, J. (Eds.) *COST Action 'Comparative Research into Current Trends in Public Sector Organization' (CRIPO)*, compendium book, forthcoming.
- Harrison, S. & Pollitt, C. (eds.) (1994) *Controlling health professionals*, Buckingham, Open University Press.
- Hartley, J. (1983) Ideology and organizational behavior, *International Studies of Management and Organization* 13(3), 7-34.
- Healthcare Commission (2009) *Investigation into Mid Staffordshire NHS Foundation Trust*, London, Healthcare Commission, March.
- Holt, D. (2008) Official statistics, public policy and trust, *Journal of the Royal Statistical Society*, A171(2), 323-346.
- Hondeghem A. & Depré R. (2005) *De Copernicushervorming in perspectief: Veranderingsmanagement in de federale overheid*. Vanden Broele, Brugge.
- Hood, C. (1990) De-Sir Humphreyfying the Westminster Model of bureaucracy: a new style of governance? *Governance* 3(2), 205-214.
- Hood, C. (1991) A public management for all seasons. *Public Administration* 69(1), 3-19.
- Hood, C. (2005) 'Public management. The word, the movement, the science', pp. 7-26 in Ferlie, E., Lynn Jnr., L. & Pollitt, C. (eds.) *The Oxford handbook of public management*, Oxford, Oxford University Press.
- Hood, C. (2011) 'Book review. The Ashgate Research Companion to New Public Management, *Governance* 24(4), 737-739.
- Hood, C. & Jackson, M. (1991) *Administrative argument*, Aldershot, Dartmouth.
- Hood, C. & Peters, G. (2004) The middle aging of New Public Management: into the age of paradox? *Journal of Public Administration Research and Theory* 14(3), 267-282.
- Independent Evaluation Group (2008) *Public sector reforms: what works and why*, Washington DC, World Bank.
- Jackson, B. (2001) *Management gurus and management fashion*, Routledge, London and New York.
- Jacobs, R. & Goddard, M. (2007). How Do Performance Indicators Add Up? An Examination of Composite Indicators in Public Services. *Public Money & Management*, 27(2), 103-110.
- James, O. (2010) *End user satisfaction as a performance measure for public services*. Paper for the National Audit Office, Department of Politics, University of Exeter, February.
- James, O. (2011) Performance measures and democracy: information effects on citizens in field and laboratory experiments, *Journal of Public Administration Research and Theory* 21(3), 399-418.
- Jeannot, G. & Guillemot D. (2010) Réformer par les outils ou par les hommes? Un bilan quantitatif de la « modernisation de la gestion de l'Etat », *Politiques et Management Public* 27(4), 73-101.
- Joss, R. & Kogan, M. (1995). *Advancing quality: Total Quality Management in the National Health Service*. Buckingham: Open University Press.
- Kaufmann, D., Kraay, A. & Mastruzzi, M. (2007) Governance matters VI: governance indicators for 1996-2006. World Bank Policy Research Working Paper 4280.
- Kearns, A. & Forrest, R. (2001) Social cohesion, social capital and the neighbourhood. *Urban Studies*, 38(12), 2125-2143.
- Kelman, S. & Friedman, J. (2009) Performance improvement and performance dysfunction: an empirical examination of the distortionary impacts of the emergency room wait-time target in the English National Health Service. *Journal of Public Administration Research and Theory* 19(4), 917-946.

- Koppenjan, J. & Klijn, E-H. (2004) *Managing uncertainties in networks* London and New York, Routledge/Taylor and Francis.
- Kuhlmann, S., Bogumil, J. & Grohs, S. (2008) Evaluating administrative modernization in German local governments: success or failure of the “New Steering Model”? *Public Administration Review*, September/October, 851-863.
- Kurki, M. (2008) *Causation in international relations; reclaiming causal analysis*. Cambridge: Cambridge University Press.
- Laforgue D. (2004) Le repli gestionnaire des inspecteurs d’académie, in *L’action publique au travail*, coll. Le travail en débats, éd. Octares, 169-176.
- Lægread, P., Roness, P. & Rubecksen, K. (2008) Performance information and performance steering: integrated system or loose coupling? In W. van Dooren & S. van de Walle (Eds), *Performance information in the public sector: how it is used*. Basingstoke: Palgrave.
- Lember, V. (2008) *New Public management reforms, contracting out and Emergency Medical Service in Estonia*, paper presented at the European Consortium for Political Research workshops in Rennes, September.
- Löffler, E. (2000) Best-practice cases reconsidered from an international perspective, *International Public Management Journal* 3, 191-204.
- Lynn, L. Jnr. (2006) *Public management: old and new*, London and New York, Routledge/Taylor and Francis.
- Macinati, M.S. (2006) Il ricorso all’ outsourcing nel Ssn: i risultati di un’indagine empirica, *Mecosan* 57, 121-144.
- Magee, H., Lucy-Davis, J. & Coulter, A. (2003) Public views on healthcare performance indicators and patient choice, *Journal of the Royal Society of Medicine* 96, 338-342.
- Makinson, J. (2000) ‘Incentives for change, rewarding performance in national government networks’, Public Services Productivity Panel, HM Treasury, London.
- Margetts, H. (2011) Experiments for public management research *Public Management Review* 13(2), 189-208.
- Marsden, D. & French, S. (1998) ‘What a performance, performance related pay in the public services’, Centre for Economic Performance, London School of Economics.
- Matelly, J-H & Mouhanna, C. (2007) *Police: des chiffres et des doutes: regard critique sur les statistiques de la délinquance*, éd. Michalon, Paris.
- McNabb, D. (2010) *Case research in public management*. Armonk, NY: M.E. Sharpe.
- McNulty, T. & Ferlie, E. (2004) Process transformation: Limitations to radical organizational change within public service organizations, *Organization Studies* 25(8), 1389-1412.
- Meyer-Sahling, J-H. (2009) *Sustainability of civil service reforms in Central and Eastern Europe five years after EU accession*, Sigma paper No.44, Paris, OECD.
- Minister for Government Policy (2011) *Open public services white paper*, Cm8145, London, The Stationary Office.
- Minvielle, E. (2006) New public management à la française: The case of regional hospital agencies, *Public Administration Review*, September/October.
- Mottram, R. (2008) *Fifteen years at the top in the UK civil service – some reflections*, Speech to MPA Capstone Event, London School of Economics, 6 May.  
(<http://www2.lse.ac.uk/study/graduate/taughtProgrammes2011/MPA/PDFs/SirRichardMottramtxt.pdf>)
- Nagyistók S. (2010) Vízitdíj, avagy egy elvetélt kísérlet, *IME IX. Évfolyam* 3. Szám 2010. Április.
- National Audit Office (2001) *Measuring the performance of government departments*, HC301, Session 2000-2001, London, National Audit Office.
- National Audit Office (2009) *Private finance projects, a paper for the Lords economic affairs committee*, October 2009, London, National Audit Office.

- Nemec, J. & Kolisnichenko, N. (2006) Market-based health care reforms in Central and Eastern Europe: lessons after ten years of change, *International Review of Administrative Sciences* 72(1), 11-26.
- Nemec, J. (2007) Performance management in the Slovak higher education system: preliminary evaluation, *Central European Journal of Public Policy*, 1(1), 64-78.
- Nemec, J. (2008) Market based health care reforms solutions and their outcomes in CEE. In Antkiewicz, S., Kalinowski, M. (eds.) *Tendencja i uwarunkowania rozwoju finansow przedsiebiorstw i finansow publicznych*. S. 338-352, Warszawa a Gdansk: CeDeWu a WSB Gdansk.
- Niskanen, W. (1973) *Bureaucracy: servant or master?* London, Institute for Economic Affairs.
- Norges Forskningsråd (2007) Resultatevaluering av sykehusreformen: *Tilgjengelighet, prioritering, effektivitet, brukermedvirkning og medbestemmelse*.
- Nunberg, B. (1999) *The state after communism: administrative transitions in Central and Eastern Europe*, Washington DC, World Bank.
- OECD (1995) *Governance in transition: public management reforms in OECD countries*. Paris: PUMA/OECD.
- OECD (2009) *Government at a glance 2009*. Paris: OECD.
- OECD (2010) *Modernising the public administration, a study on Italy*. Paris: OECD.
- OECD (2011) *Government at a glance 2011*. Paris: OECD.
- Office of Public Services Reform (2002) *Better government services: executive agencies in the 21<sup>st</sup> century*. (<http://www.civilservice.gov.uk/agencies>).
- Ongaro, E. (2009) *Public management reform and modernization: trajectories of administrative change in Italy, France, Greece, Portugal and Spain*, Cheltenham and Northampton, Mass., Edward Elgar.
- Osborne, D. & Gaebler, T. (1992) *Reinventing government: how the entrepreneurial spirit is transforming the public sector*. New York, NY: Addison Wesley Publishing Company.
- Osborne, S. (ed.) (2010) *New public governance: emerging perspectives on the theory and practice of public governance*, London and New York, Routledge/Taylor and Francis.
- PA Consulting Group (2002) *Evaluation of the Strategic Management Initiative*. Report presented to the Department of the Taoiseach, Dublin, PA Knowledge Ltd.
- Painter, M. & Peters, G.B. (eds.) (2010) *Traditions and public administration*, Basingstoke, Palgrave/Macmillan.
- Pawson, R. (2002) Evidence-based policy: the promise of “realist synthesis. *Evaluation* 8(3), 340-358.
- Pawson, R., Greenhalgh, T., Harvey, G. & Walshe, K. (2005). Realist review – a new method of systematic review designed for complex policy interventions. *Journal of Health Service Research Policy*, 10(11), S1.21-S1.34.
- Pollitt, C. (1990) *Managerialism and the public services*, Oxford, Blackwell.
- Pollitt, C. (1995) Justification by works or by faith? Evaluating the New Public. *Evaluation* 1(2), 133-154.
- Pollitt, C. (2003) *The essential public manager*. Buckingham: Open University Press/McGraw Hill.
- Pollitt, C. (2007) New Labour’s re-disorganization. A cautionary tale’, *Public Management Review* 9(4), 529-543.
- Pollitt, C. (2010) Simply the best? The international benchmarking of reform and good governance, pp. 91-113 in J.Pierre and P.Ingraham (eds.) *Comparative administrative change and reform: lessons learned*, Montreal and Kingston, McGill-Queen’s University Press.
- Pollitt, C. & Bouckaert, G. (Eds) (1995) *Quality improvement in public services: concepts, cases and commentary*. London: Sage.
- Pollitt, C., Birchall, J. & Putman, K. (1998a) *Decentralising public service management*. Basingstoke: Macmillan.
- Pollitt, C., Keen, J., Packwood, T. & Roberts, S. (1998b) *The Transforming Healthcare Delivery programme at King's Healthcare: an evaluation*, Uxbridge, Centre for the Evaluation of Public Policy and Practice, Brunel University, February.
- Pollitt, C. & Bouckaert, G. (2000) *Public management reform: a comparative analysis*. Oxford: Oxford University Press.

- Pollitt, C., Talbot, C., Caulfield, J. & Smullen, A. (2004) *Agencies: how governments do things through semi-autonomous organizations*, Basingstoke, Palgrave/Macmillan.
- Pollitt, C. & Bouckaert, G. (2009) *Continuity and change in public policy and management*. Cheltenham: Edward Elgar.
- Pollitt, C., Harrison, S., Dowswell, G., Jerak-Zuiderent, S. & Bal, R. (2010) Performance regimes in health care: institutions, critical junctures and the logic of escalation in England and the Netherlands, *Evaluation* 16:1, 13-29.
- Pollitt, C. & Bouckaert, G. (2011) *Public management reform: a comparative analysis: New Public management, Governance and the Neo-Weberian State*. Oxford: Oxford University Press.
- Pollitt, C. & Hupe, P. (2011) Talking about government. The role of magic concepts, *Public Management Review* 13(5), 641-658.
- Pollitt, C. (2011) Moderation in all things: international comparisons of governance quality. *Financial Accountability and Management* 27(4), 437-457.
- Prime Minister and the Minister for the Cabinet Office (1999) *Modernising government*, Cm. 413, London, The Stationery Office.
- Promberger, K. (2008) *Evaluation der Verwaltungsmodernisierung in Österreich am Beispiel des Pilotprojekts Flexibilisierungsklausel*, Jubiläumsfondsprojekt Nr. 11998, Innsbruck, Juli 2008.
- Propper, C., Sutton, M., Whitnall, C. & Windmeijer, F. (2008) *Incentives and targets in hospital care: evidence from a natural experiment*. Bath, CMPO Working Paper 9.
- Randma-Liiv, T. (2005) Performance management in transitional administration: the introduction of pay for performance in the Estonian civil service, *Journal of Comparative Policy Analysis Research and Practice* 7(1), 95-115.
- Roberts, A. (1997) Performance-based organizations: Assessing the Gore Plan, *Public Administration Review* 57(6), 465-478.
- Roberts, A. (2010) *The logic of discipline: global capitalism and the architecture of government*. Oxford: Oxford University Press.
- Rossmann, B. (2001) Ausgliederungen – Kein ersatz für eine reform des öffentlichen sektors, *Wiso*, 3, Linz, 139-157.
- Sahlin-Andersson, K. & Engwall, L. (2002) *The expansion of management knowledge: carriers, flows and sources* Stanford CA, Stanford University Press.
- Saint-Martin, D. (2000) *Building the managerialist state: consultants and the politics of public sector reform in comparative perspective*, Oxford, Oxford University Press.
- Societatea Academică Română (2010) Pentru o autonomie responsabilă, expansiunea universităților în România, SAR Policy Brief No. 46, București, Iulie.
- State Audit Office (2002) Utilization of Payroll Funds Allocated for the Payment of Performance Wages, State Audit Office, Tallin, Estonia.
- Sundström, G. (2006) Management by results: its origins and development in the case of the Swedish state *International Public Management Journal* 9:4, 399-427.
- Șandor, D. S & Tripon, C. R. (2008) Citizens' perceptions about public administration reforms in Romania, *Transylvanian Review of Administrative Sciences* 24E, 111-122.
- Torres, L. & Pina, V. (2004) Reshaping public administration: the Spanish experience compared to the UK. *Public Administration* 82(2), 445-464.
- Torres, L., Pina, V. & Royo, S. (2005) E-government and the transformation of public administrations in EU countries: beyond NPM or just a second wave? *Online Information Review* 29(5), 531-553.
- Tsoukas, H. (2005) *Complex knowledge: studies in organizational epistemology*. Oxford: Oxford University Press.

- Van de Walle, S., Van Roosbroek, S. & Bouckaert, G. (2008) Trust in the public sector: is there any evidence for a long term decline? *International Review of Administrative Sciences* 74(1), 47-64.
- Verheijen, T. (2007) *Administrative capacity in the new EU member states: the limits of innovation?* World Bank Working Paper No.115, Washington DC.
- Verheijen, T. & Dobrolyubova, Y. (2007) Performance management in the Baltic States and Russia: success against the odds? *International Review of Administrative Sciences* 73(2), 205-215.
- Verhoest, K. (2005) Effects of autonomy, performance, contracting and competition on the performance of a public agency: a case study, *Policy Studies Journal* 33:2, 235-258.
- Walker, D. (2011) 'Cameron's faith in the market should have been shaken by Southern Cross', *Guardian Professional*, 18 July (<http://www.guardian.co.uk/public-leaders-network/2011/jul/18>, accessed 21 July 2011).
- Wilson, D. (2011) Comparative analysis in public management, *Public Management Review* 13(2), 293-308.
- Wilson, D. & Piebalga, A. (2008) Performance measures, ranking and parental choice: an analysis of the English school league tables, *International Public Management Journal* 11:3, 344-366.
- World Bank (2001) Salary supplements and bonuses in revenue departments, Final report, World Bank.
- Yesilkagit, A. and De Vries, J. (2002) The unanticipated consequences of decentralization and reinvention: the case of the Province of South Holland, *International Review of Administrative Sciences* 68(4), 579-597.

## Acknowledgements

We wish to acknowledge our main – and indispensable – collaborators, consisting of the teams at the other COCOPS partner universities: Erasmus University Rotterdam, the Hertie School of Governance, the University of Bergen, Bocconi University, the University of Cantabria, Cardiff University, Corvinus University Budapest, the University of Exeter, Tallinn University of Technology and the Centre National de la Recherche Scientifique (CNRS – Paris). A series of public administration and management scholars and researchers from within and outside the COCOPS network have given us useful feedback on previous drafts of this report. We list them here while apologizing to anyone we may, inadvertently, have missed: Rhys Andrews, Philippe Bezes, Geert Bouckaert, Tom Christensen, György Hajnal, Gerhard Hammerschmid, Christopher Hood, Oliver James, Riin Kruusenberg, Per Læg Reid, Edoardo Ongaro, Tiina Randma-Liiv, Lise H. Rykkja, Steven Van de Walle, Koen Verhoest, Kai Wegrich. We would also like to thank the participants in the Permanent Study Group II, Performance in the public sector, for their valuable comments during the European Group of Public Administration (EGPA) conference, 7-10 September 2011 in Bucharest. Stijn Schoors of the K.U.Leuven's central IT department provided highly competent technical support in setting up and maintaining the SharePoint system.

We also wish to thank a number of individuals and institutions who voluntarily offered us material and advice, although they were not formal COCOPS partners. These were:

Dan Apăteanu, Babeş-Bolyai University, Romania  
Yuliyán Atanasov Gospodinov, Tsenov Academy of Economics, Bulgaria  
Maria Asensio, National Institute of Administration, Portugal  
Dario Barbieri, Bocconi University, Italy  
Dana Bajusova, Economics University of Bratislava, Slovakia  
Ioana Ciucanu, National School of Political Studies and Public Administration, Romania  
Anna Czarczynska, Leona Kozminskiego Academy Warzaw, Poland  
Iwona Grabowska, Collegium of World Economy, Warzaw School of Economics, Poland  
Tomas Jacko, Comeniun University Bratislava, Slovakia  
Milena Krasimirova Vladimirova, Medical University of Varna, Bulgaria  
Ivan Kopric, Zagreb Law Faculty, Croatia  
Georgi Dimireov Manliev, Technical University of Sofia, Bulgaria  
Monica Marin, Research Institute for Quality of Life, Romania  
Lucica Matei, National School of Political Studies and Public Administration, Romania  
Miruna Mazurencu-Marinescu, Academy of Economic Studies, Romania  
Marco Meneguzzo, University of Rome Tor Vergata, Italy  
Michal Mierzwa, Krajowa School of Public Administration Warzaw, Poland  
Beata Mikusova Merickova, Matej Bel University Banska Bystrica, Slovakia  
Cristina Mititelu, University of Rome Tor Vergata, Italy  
Vitalis Nakrošis, Vilnius University, Lithuania  
Juraj Nemeč, Matej Bel University Banska Bystrica, Slovakia  
Goranka Lalic Novak, Social Sciences Polytechnic Zagreb, Croatia  
David Spacek, Masaryk University, Czech Republic  
Robert Szczepankowski, Stanislaw Staszic College of Public Administration, Bialystok,  
Poland  
Piotr Szreniawski, Maria Curie Sklodowska University, Lublin, Poland  
Ioan Toderaş, Centre for Promoting Research and Innovation, Romania  
Adelina Dumitrescu, National School of Political Studies and Public Administration,  
Romania  
Agne Vabamae, Tallin University of Technology, Estonia  
Mădălina Voican-Avram, University of Craiova, Romania  
Dominika Wojtowicz, Leona Kozminskiego Academy Warzaw, Poland  
Margarita Yordanova Bogdanova, Tsenov Academy of Economics, Bulgaria

