

PDF hosted at the Radboud Repository of the Radboud University Nijmegen

The following full text is an author's version which may differ from the publisher's version.

For additional information about this publication click this link.

<http://hdl.handle.net/2066/111791>

Please be advised that this information was generated on 2021-05-09 and may be subject to change.

Ranking and benchmarking: The political logic of new regulatory instruments in the fields of gender equality and anti-corruption

Anna van der Vleuten and Mieke Verloo

New regulatory instruments such as benchmarking, ranking and best practices have given new meaning to the old politics of reputation and changed the standards by which performance is judged. In this article, we investigate the power-based mechanisms that operate between reputation, ranking, benchmarking and performance. Analysing two cases – gender equality policies at EU level and the fight against corruption at the global level – we show that there are valid reasons to doubt the contribution of ranking and benchmarking to improving policy making. The underlying political dynamics even result in adverse effects on performance in terms of legitimacy and effectiveness.

A political discussion of new regulatory instruments

Since the 1980s, New Public Management (NPM) and its associated doctrines of public accountability and organisational best practice have given birth to a large array of new regulatory instruments meant to improve the performance and the political accountability of public administration systems (de Vries, 2005). Some argue that these innovations have made ‘public management in most OECD [Organisation for Economic Co-operation and Development] countries more efficient, more transparent and customer-oriented, more flexible, and more focused on performance’ (Haque, 2007: 181). Others blame NPM for its adverse effects on the quality of service delivery and the downsizing of public employment. Discussions of NPM have often been technical, aiming to improve the quality of the instruments, or ideological, criticising its market-centric neoliberal assumptions (Haque, 2007).

What is lacking is a political discussion of NPM. Caporaso and Wittenbrinck (2006: 476) are right in pointing out that this new mode of governance ‘is itself a political act’, but they limit their discussion of the political dimension to the shift from majority voting to consensus. Of course, there is more to the political nature of NPM than voting rules. We argue that an analysis of the power-based logic underlying these instruments will enable us to better understand why they often produce perverse effects and undermine the quality and transparency of policy making instead of improving it. We focus on two sets of instruments: policy steering through ‘pioneer standard setting’ (benchmarking, targets and pilot projects) and enforcement through ‘relative standard setting’ (ranking of performance and peer review).

These instruments are widely applied not only in the national context, transforming in the 1980s for instance the American and the British civil services (Majone, 1997), but also internationally. We will explore their application in the international arena, where political authority is decentralised and accordingly the

Key words: reputation • benchmarking • gender equality • corruption • new regulatory tools

power-based logic is revealed even more clearly than elsewhere. We have selected two cases, gender equality and the fight against corruption, which deal with European policy making and global policy making respectively. Since they concern different policy arenas, we will be able to show how these new regulatory instruments lead to similar consequences in spite of institutional differences. The choice of policy issues has been guided by our assumption that the perverse effects are stronger if the issue is perceived to be more ideological than technical. In that respect, gender equality and anti-corruption constitute domains *par excellence*, as good performance in these fields contributes to the reputation of a state as a 'good guy'.

Old and new regulatory instruments and their power-based logics

New regulatory instruments differ from old-style politics in many respects, of which three are particularly relevant here. They concern the definition or representation of the problem, the choice of the problem-solving strategy and the enforcement of the solution.

The 'old' instruments of international governance were clearly power based as the definition of the problem as well as the choice of problem-solving strategies were made through political bargains 'negotiated tooth and claw by groups with conflicting objectives, accompanied of course by winners and losers' (Caporaso and Wittenbrinck, 2006: 476). Enforcement was obtained through the 'politics of reputation'. States value their reputation; it is part of their power position with regard to domestic society and to other states. Reputation is based primarily on the quality of past performance. If a state has built up trust, it can convince others of the quality of its ideas more easily and will be more successful in promoting its model. Conversely, 'being perceived as a cheater and free-rider' undermines the state's bargaining power (Börzel, 2003: 203) and the attractiveness of its practices. A state's interest in a good reputation thus contributes to investments in better performance and plays a role in explaining state compliance with international standards, particularly regarding issues such as gender equality and corruption that are linked to its identity as a 'good state' (van der Vleuten, 2005).

If a state is unwilling to implement an international standard in these domains, it will prefer its non-compliance to go unnoticed. Accordingly, international organisations have set up monitoring systems, increasing the probability that non-compliance is discovered. In the European Union (EU), the Commission is entitled to investigate whether a state rhetoric corresponds with the facts, and to sanction non-compliance by bringing a state before the European Court of Justice. Member states strongly dislike such procedures, since the Commission usually wins the case and Court rulings receive a great deal of publicity, especially when financial sanctions are imposed. The involvement of the Court implies that an absolute standard is applied, meaning that a third party provides the final interpretation of a rule that is applicable to all member states, business and citizens. Also, in its interpretations, the Court takes into account the shared understandings of member states as regards the issue at hand. The strengths of these 'old school' politics of reputation thus are constituted by the reliance on absolute standards that are authoritatively interpreted and enforced by a third party, transparency, and the high reputation costs of non-compliance. Other

international organisations do not have such strong tools for authoritative standard setting and enforcement at their disposal, however. Also, the system suffers from another weakness: the judgement of performance is strictly formal, assessing the legal implementation of a standard, which often does not adequately reflect the *de facto* attainment of the policy goal contained in the standard (Lombardo, 2003).

New regulatory instruments aim to repair such weaknesses. They tend to be presented 'in a positive normative light', suggesting that the processes of problem definition and problem solving represent a break with authoritative forms of policy making and are characterised by deliberation and the bottom-up involvement of social actors (Caporaso and Wittenbrinck, 2006: 476). Let us take a closer look at some new instruments.

Best practice is one such new instrument. Best practices are assumed to offer the best solution to a problem, and are presented as a model to be copied in other states or situations. This method does not result in an absolute standard, but in what we have dubbed a 'pioneer standard'. A pioneer standard is a context-specific way of coping with a problem that is dressed up as an absolute standard without specifying the scope conditions that facilitated its success. Although it is no absolute standard, all parties involved are urged to follow the pioneer and apply its methods. However, the effectiveness of this urge is limited, since non-conformity cannot lead to coercive measures and shaming will be of limited use, as the development of different practices does not necessarily preclude solving the problem for which the best practice was recommended. Therefore, the links between best practice and reputation are not straightforward. The connection may even be reversed, since producing best practices not only contributes to good reputations, but good reputations also produce best practices. In sociology, this is known as the 'Matthew effect', which 'consists in the accruing of greater increments of recognition for particular scientific contributions to scientists of considerable repute and the withholding of such recognition from scientists who have not yet made their mark' (Merton, 1968: 58). In fact, to deck out a new strategy, officials will collect best practices before any practice even exists. If they base their selection on proposals by government representatives, ideas proposed by states with a good reputation in the field are more likely to receive the best practice label than ideas proposed by notorious laggards.

Ranking is an instrument meant to improve compliance with policy goals. It works through the establishment of indicators. Instead of laying down a legal standard, policy goals are disaggregated in quantifiable objectives, and based on their score on such indicators, states may then be ranked and classified as pioneers or laggards. This means that a relative standard is applied. Policy makers and experts decide which aspects of performance will be measured. As different states will be at an advantage or disadvantage depending on the chosen ways of measuring, it then matters which procedures are adopted to select indicators, and who has the power to decide which data are to be produced. Rather than developing the indicators in a transparent democratic debate, there is a tendency to leave this to experts, based on a strong 'faith in the power of expertise' (Majone, 1997: 152). For the sake of comparison, the ranking objectives tend to be one-dimensionally quantified, and indicators selected on the basis of the availability of data. This may lead to the 'stretching and bending' of policy objectives to fit the available data (Lombardo et al, 2009). Another pitfall is the phenomenon of 'reactivity', in which actors strategically play into indicators

to raise their scores, regardless of whether they actually perform well (Espeland and Sauder, 2007).

In short, there are no standards resulting from negotiations between state representatives and interpretations by judges, but only relative and pioneer standards resulting from technocratic exercises. Experts and the availability of data will define the problem. This of course casts doubts as regards the legitimacy of the problem definition and raises questions as to who can be held accountable. The next sections on European gender equality policies and global anti-corruption policies will highlight the political dynamics involved in new regulatory instruments, distorting the link between performance and reputation.

Promoting gender equality in the European arena

Gender equality is a domain in which, for more than 50 years, the EU has strongly influenced national policies, especially through legislation that focused on the labour market. The European Commission and the Court monitored and enforced compliance, sometimes succeeding in forcing unwilling states to implement the European rules (van der Vleuten, 2007). In the 1990s, a shift took place from legal instruments to instruments such as ranking and best practices under the label of the Open Method of Coordination (OMC) (Héritier, 2003).

The OMC was formally introduced in 1997 in the Treaty of Amsterdam as part of the European Employment Strategy (EES). This strategy was developed in order to tackle high unemployment across Europe without transferring any real policy-making powers to the European level (Mosher and Trubek, 2003). It asks European government leaders to approve policy guidelines, for example on equal opportunity policies to promote gender equality. The guidelines have to be implemented by the OMC, a method that relies on soft mechanisms such as peer review and benchmarking. It aims to promote mutual learning and foster the generation and diffusion of new ideas and practices across Europe. To that effect, the European Commission identifies successful member states and incorporates their practices in the guidelines. After the Council of Ministers has approved the guidelines and set targets, each member state draws up a National Action Plan, presenting its strategies to reach the targets. Progress is monitored through peer review. Each year the Commission examines compliance with the guidelines and proposes recommendations to individual member states. The assessments and recommendations are included in the Joint Employment Report, which leads to a renewed identification of best practices, and to a new start of the cycle (Mosher and Trubek, 2003). There are fierce debates as to whether or not the European Employment Strategy is delivering on its promises (Héritier, 2003; Pfister, 2008). Here we will focus on the consequences of the use of indicators and benchmarking for performance in the field of gender equality.

Defining the problem

The EES has resulted in an ongoing production of member state rankings on employment-related issues. The most important targets formulated in the domain of gender equality are those concerning women's employment and unemployment

rates, gender gaps in pay, and the provision of childcare. Although measuring the performance on the chosen indicators seems to 'prove' which states are doing well, and which states are underperforming, the method suffers from several biases.

States have different ways to frame gender inequality as a policy problem and define gender equality as a goal (Bacchi, 1999; Verloo, 2007), but these differences tend to disappear as the definitions are made dependent on the availability of data. This means that we find more indices on gender inequality in the public sector than in the private domain, more indicators on numbers of women than on empowerment of women in the sense of giving voice to non-hegemonic actors in a society (Verloo, 2005). As a result, the indices, although they aim to measure progress towards gender equality, measure simple inclusion instead of structural progress.

The use of indices contributes to a 'uni-dimensionalisation' of the gender inequality *problématique* in two ways: the complex interrelationships between gender inequalities in citizenship, labour and intimacy are not captured anywhere (Verloo, 2007), and women are treated as a unitary category rather than as a category also constituted by its intersections with other inequalities (class, race/ethnicity, sexual orientation, age, physical ability). Limiting the quality of performance to targets that are measured by mono-dimensional indices leads national authorities to focus on the production of documents on these indices. This redefines the problem of gender inequality as a technical problem in the sense of missing data and expertise, and incentivises improving the quality of the documents on gender equality instead of improving gender equality itself. This is one form of 'reactivity' that is counterproductive to performance (Espeland and Sauder, 2007).

Limiting the assessment of performance to indices and targets may lead to misleading conclusions. In one case, the EU praised Greece for closing the gender employment gap, disregarding the fact that Greece achieved this thanks to a drop in the male employment rate (Rubery, 2005). Similarly, changes in the segregation index were interpreted as progress in desegregating labour markets, without these changes being decomposed into effects related to, respectively, changes in the labour force and changing gender shares within occupational categories (Rubery, 2005).

Although there is good reason to complain about the lack of indices (Luxembourg Presidency Report, 2005), we argue that the main problem is not the mere number of indices, but the implicit and explicit selection mechanisms that underlie the choice of indices. Certain fields of gender equality are given more importance than others. The most developed indices are found in the economic field, encompassing aspects of education, social inclusion and poverty (Walby, 2005), while the fields of citizenship and gender-based violence are largely ignored. This bias not only influences the rankings that result from the indices, but also the politics of reputation. In 2000, for instance, the traditional laggard Austria developed a law on domestic violence based on an innovative model of state intervention in this field. Not until six years later, the Netherlands – an alleged pioneer – introduced similar legislation. Still, introducing its law early has not changed Austria's reputation as a laggard in gender equality policies. This may be connected to the fact that there are no targets or indices on gender violence.

A classic 'technocratic' complaint would be that the EES indices are not well constructed. Indeed, different indices can produce different rankings. An index based on the equal opportunity indices from the EES would put the usual laggards

Portugal and Belgium ahead of gender pioneers Denmark and Sweden, and would put the usually high-ranking Finland in ninth place (Plantenga et al, 2003: 21). A different index, based on a set of indices that includes aspects of political and private contexts, presents a ranking that fits the classical expectations of 'good' gender equality states, listing Sweden, Denmark and Finland in the top three (Plantenga et al, 2003: 37). The point here is not to argue about the respective qualities of indices, but to accentuate that every choice for one over the other has substantial implications for states' ranking, as well as for policy recommendations to improve one's performance.

Relative standard setting, pioneer standard setting and compliance

The EES targets do not constitute an absolute standard based on a consensus over which situation is good or bad, the way in which legal instruments define certain practices as illegal. They constitute relative standards indicating that providing childcare for 33% of the youngest children is better than, say, 21%, but this does not mean that 33% is in itself a good or 21% a bad situation. Therefore, these targets are arbitrary standards for judging performance. How then to adequately judge state behaviour? In addition, these are not *de facto* targets for everyone, as some member states are outperforming them from the start. These member states are not compelled to undertake further action, as they do not face an absolute standard with which to comply. Rönnblom (2004: 2) notes how Sweden proudly presents itself as the 'most gender equal country in the world', thereby legitimising its inaction. Relative targets may legitimise inaction not only by the leaders but also by the laggards. As Mósesdóttir (2006: 53) argues, ranking runs the risk of 'making those countries that outperform the others the unachievable standard for member countries falling behind. Economic, political and social contexts cannot be copied over night'. 'Laggard' Italy, for instance, reports that the EES guidelines are ineffective because they are not appropriate for the Italian labour market (Mosher and Trubek, 2003). As a result, the annual EES scoreboard shows persistent classes of leaders (the Scandinavian countries) and laggards (the Mediterranean countries) without the latter catching up with the former.

Another building block of the EES is the production of best practices. Member states are 'encouraged to benchmark their performance against the best performer in the Union' (Héritier, 2003: 117). This implies that one national solution is considered the best. Yet, the selection of the best practice is a political and not a technical exercise, as it does not consist of questioning *all* practices and investigating which practice suits which context best. Networks of gender experts assist the European Commission in the assessment of member state performance. They are a source of 'labelling' best practices and they create pioneers in the process. In their review of all the performed European gender impact assessments (GIAs), Rubery and Fagan (2000) position the Netherlands as an early starter. They then incorporate the Dutch conceptual framework as an academic model for what they recommend as a required further conceptual and practical elaboration of the GIA methodology. In practice, though, the Dutch GIA methodology has proven to be difficult to export because it is tailored to Dutch data and Dutch framing of gender equality

(Roggeband and Verloo, 2006). This shows that best practices have little value if they are not contextualised.

Compliance through policy learning and shaming

The process of peer review presumes that member states perceive themselves as a collective that is eager to learn. However, member states often have not been eager to call into question their national practices. It has been the enforcement procedures involving the Commission and the Court that have forced them to do so more than once. In the 1980s, for instance, the Dutch and British governments had to revise their breadwinner-centred systems of social security thoroughly in order to comply with a European directive that ruled out discrimination in social security (van der Vleuten, 2007). If instead of being taken to court they had been asked to imitate the Danish best practice of an individualised social security system, Dutch and British representatives probably would have explained that in their national context the breadwinner system was the best practice to ensure an equitable family income. The domestic political and economic costs involved in policy change would have severely limited their presumed 'eagerness to learn'. Why would these national officials and politicians want to be driven by mimesis when facing domestic opposition?

The only other instrument available to make member states comply is to designate them as 'villains' or 'heroes' (Barysch et al, 2008). However, as we have argued already, a top ranking does not always mean that the state has behaved heroically, and a low-ranking state has not necessarily behaved like a villain, since member states may have done their best in spite of unfavourable conditions or may have rested on their laurels in spite of favourable conditions. For an underperforming state, the reputational costs of the EES remain limited, as long as it is all about relative standards without a third party authoritatively assessing performance as compliance or non-compliance. How would shaming bring gender-traditional Malta, where the gender gap in employment rates is as high as 37% (measured in full-time equivalent), to join Finland and Latvia, where the gap is less than 8% (EC, 2009)?

One may wonder whether the old system was effective. The answer is to a certain extent equivocal. France, for instance, cherishes its reputation as a social role model. From the very beginnings of the European integration process, it has been vocal about gender equality (van der Vleuten, 2007). Just when France occupied the EU presidency, in spring 2000, the European Court was on the verge of condemning it to a daily fine of 142,425 euros if it failed to lift the ban on night work for women. The French government succeeded in convincing its utterly divided Parliament of the necessity to act immediately by pointing at the consequences for its reputation. To be the first member state ever to be condemned to a penalty payment by the European Court over gender equality policy was unbearable for a would-be champion of equality like France (van der Vleuten, 2007). As a result, France complied by lifting the ban on night work. In fact, this is a telling example of the strengths and weaknesses of the legal system: compliance with the standard was obtained and legal discrimination eliminated, but the labour conditions of many French women deteriorated as they now had to accept night shifts.

In their review of the EES, Mosher and Trubek (2003: 78) found that some policy changes 'came about because exchange of information and deliberation within the

EES process brought new ideas to the fore'. For controversial policies, however, the political mandate remains decisive (de la Porte and Nanz, 2004). In addition, the annual peer review process does not seem to encourage learning in the sense of developing shared discourses and diffusing knowledge, since less than an hour is allocated for the presentation and discussion of each National Action Plan. Some member states treat the presentation of the National Action Plans as a bureaucratic reporting task instead of an opportunity for reflection (Mosher and Trubek, 2003). As the monitoring and the ensuing naming and shaming do not focus on the process but on the outcome, not the learning process but the production of the 'correct' information becomes the key objective. Member states only have an interest in producing data instead of adopting an open attitude towards each others' practices, turning the OMC into a 'mere beauty contest' or 'an exercise in statistics' (Borrás and Jacobsson, 2004: 195).

Transparency and accountability

Until the 1990s, the public accountability of EU politics was contested, because decision-making powers were concentrated in the hands of a Council of Ministers that deliberated behind closed doors. The introduction of the co-decision procedure in 1992, which gave the directly elected European Parliament a formal say, has strongly improved political accountability. The European Court and the national courts jointly assured judicial accountability. Although the increase of transparency and accountability was another objective of NPM, the introduction of new regulatory instruments has actually reduced accountability, because the European Court is not involved in the interpretation of the 'soft law' standards that result from the OMC (Borrás and Jacobsson, 2004). Member states do not have the option to challenge their ranking in court. Political accountability is problematic as well, as the European Parliament is only involved in strict advisory terms. At the national level, there is no public debate about the EES; we see only sparse coverage in the national media, parliamentary debates have been quite low key, also due to a lack of knowledge of the OMC among parliamentarians, and the participation of civil society organisations is very limited (de la Porte and Nanz, 2004).

Fighting corruption in the global arena

Compared with the EU, global international organisations lack the coercive tools to effect policy change. Of course, there are international tribunals that act as authoritative third-party standard setters, but they depend on states for enforcement. The observed shift to instruments such as ranking and benchmarking can be seen as one way for international organisations to deal with this enforcement problem. Also, these instruments seem to suit the normative power of international organisations based on impartiality and technocratic expertise (Joachim et al, 2008). A field in which ranking and benchmarking have proliferated is the promotion of good governance, including the fight against corruption by international organisations such as the International Monetary Fund, the World Bank and the OECD. They use corruption indexes and ranking as policy instruments. We argue that these instruments are more political than they might seem to be.

Nowadays, the fight against corruption ranks high on the global political agenda. During the Cold War, geopolitical motives strongly influenced decisions concerning loans by international financial institutions and bilateral financial aid. Since the fall of the Berlin Wall, other considerations such as accountability, transparency and legitimacy of government in the receiving countries have become more relevant to many donors (Lambsdorff, 1999). The careful assessment of the quality of governance has thus become of utmost importance, as it influences the attractiveness of a country for public and private financial institutions. The Corruption Perception Index (CPI) is one of the instruments used worldwide to rank countries in terms of the extent to which corruption is perceived to exist among public officials and politicians. It is published yearly by Transparency International (TI), a global civil society organisation based in Berlin.

The OECD praises the CPI as a useful tool that has ‘provided the basis for substantive anti-corruption reforms in a number of countries’ (Clarke, 2000). The idea behind the index is to stigmatise countries in the lower regions of the ranking, and discourage multinational corporations and aid-granting agencies to invest in these countries. This makes it a potentially far more powerful tool than the gender equality rankings discussed in the previous section, because a bad score on the CPI can have negative financial consequences. Hence, it matters tremendously how such indices are constructed. Despite the praise of the OECD, the CPI suffers from several weaknesses.

Defining the problem and setting the standard

An impartial assessment of corruption would require a definition of corrupt practices that focuses on practices harming the quality of governance, and that takes into account institutional differences between countries. The CPI defines corruption as ‘the abuse of public office for private gain’ (Clarke, 2000). This standard suffers from two biases, which both deflect damage to the reputation of wealthy OECD countries and undermine the reputation of poor and non-Western countries. First, it focuses on those forms of corruption that predominately occur in countries characterised by a high degree of state intervention in the economy. Only illegal private-to-public transactions are defined as corruption. The policies of so-called tax havens, which have deliberately established a tax structure and financial secrecy to take advantage of the worldwide demand for tax avoidance opportunities, are not identified as corrupt. These states harm the financial interests of other states, yet are not considered to be corrupt. This implies that those practices are singled out as corrupt that take place predominately in non-Western and poor countries, while practices in Western and wealthy countries (Singapore, Switzerland) are not labelled as corrupt. TI’s definition of corruption thus both reflects and strengthens unequal power relations.

Second, the CPI only shows the receiving end (public authorities), while it leaves out the bribe payers (notably companies from leading industrial countries). In 1999, TI therefore established a new ranking, the Bribe Payers Index, assessing the propensity of companies to bribe abroad. It reveals poor compliance with the OECD Anti-Bribery Convention by companies based in OECD countries. Striking are the poor rankings of France and Italy (Transparency International, 2006). The

Bribe Payers Index thus seems to be an improvement; however, the score of a country in this index and its score in the CPI are not combined into a single ranking on corruption. As such, the Bribe Payers Index currently does not have the same potential material consequences as the CPI.

Furthermore, the TI ranking may reflect reputation rather than performance. This is due to the data collection method; as indicated by its name – Corruption Perceptions Index –, the index is not established by listing corrupt practices directly, but indirectly, based on the perception of corruption by a set of correspondents. In effect, the index is a ‘poll of polls’. It compiles the results of 17 surveys among experts and business elite. These surveys asked them to assess the extent of corruption in the countries where their company is active. These data are then generalised to the country as a whole. While this use of as many different possible sources may give an impression of democratic standard setting, we argue that appearances are misleading. Even if all surveys involve a mix of residents and non-residents of the assessed country, all respondents share a predominately Western background and work in the Western business world, thus representing a specific elite. Their perception of corruption may be biased, reflecting their culturally coloured definition of what constitutes corrupt practices. A more heterogeneous group of respondents in terms of level of education, gender, and professional and cultural background might perceive the level of corruption in a country differently (van Hulst, 2007).

The subjective character of the specific collected data is problematic as well. Clearly, corruption assessment lacks objective measurements (such as financial audits and court cases), but not all subjective measurements are problematic to the same extent. Perception-based indicators may gain in accuracy if they ask specific questions linked to real experiences instead of vague and general questions about the level of corruption in a certain country. Yet, only one organisation asks the respondents directly about their own experiences of corruption (Fjeldstad and Andvig, 2001), all other surveys used by TI consist of general questionnaires. This increases the probability that they measure the perception of reputation instead of the perception of corruption.

The variance in scores confirms the measurements’ subjectivity. Different surveys produce different results for the same country, ranging as wide as from 4.9 (corrupt) to 9.2 (almost clean) for the United States (US) in 2003.¹ TI calculates the average to establish the country score. This resulted in a US score of 7.5, ranking 18th instead of 7th immediately behind Sweden (9.3) or sharing the 39th place with Tunisia (4.9) (Transparency International, 2007). Why would the average rather than the lower or the higher of these scores provide us with a correct representation of a country’s level of corruption? As Urra (2007: 6) points out, ‘complex statistical constructions of modern aggregate indicators can easily create an illusion of quantitative sophistication that leads to interpret them as actual corruption indicators’, but they still remain subjective indicators of perceived corruption.²

Compliance through shaming

The TI index clearly exposes the shortcomings of rankings based on relative standard setting as a tool to improve performance through naming and shaming. In fact, it is unclear what the meaning of any ranking on the index is. Ranking 18th does not

mean that the country is twice as corrupt as number nine. A lower ranking than the previous year does not necessarily imply that the level of corruption has increased. In 2002, the United Kingdom (UK) ranked 10th, in 2003 it ranked 13th. Had it done worse? Perhaps not: it had scored the same in both years (8.7). If any change takes place, it is not clear when it happened, since respondents' views are based on impressions that are not necessarily limited to the calendar year (Fjeldstad and Andvig, 2001: 32). The addition of new countries to the index changes rankings without reflecting absolute change. In 1995, Indonesia ranked 41st (out of 41 countries) and in 2006 it ranked 130th (out of 163), but its score increased from 1.9 to 2.4. Of course, careful analysis shows whether this is a case of genuine improvement or not, but the point is that the results are often published without background information. Governments, opposition and media usually present changes in rankings to prove their respective cases and without applying the necessary caution.³ In the end, reputational costs are limited since a change in ranking can always be interpreted in more than one way.

It is doubtful whether this index helps to improve performance. It has been published since 1995. Over the years, there has been no significant improvement in countries' performance nor have there been any significant setbacks. The traditional high-flyers top the list (Scandinavian countries, New Zealand, Singapore) and the usual suspects fill the bottom league (African states, Haiti, Iraq, Myanmar). The distance between the lowest and the highest scores has not diminished. This might lead to the conclusion that the corruption level in individual countries changes very slowly over time. However, the impression of stagnation may also be due to methodological weaknesses linked to the previously mentioned subjective character of definition and data: the reputation of the CPI may influence people's perceptions of the corruption level in a country, and thus 'the most recent ranking may be highly dependent on previous rankings' (Fjeldstad and Andvig, 2001: 32). We suspect that a state's reputation immunises against change for better or worse, as it may be 'easier' to report corruption in countries known to have some, while reporting corruption in countries that are 'known' to be clean may have a higher threshold, calling for substantial proof. Tanzania's poor ranking, for instance, did not improve in spite of the anti-corruption measures taken by President Mkapa, 'Mr Clean'. The high rankings of France and Germany did not suffer from financial scandals involving high-ranking politicians including the President and the Chancellor respectively.⁴ Singapore and Switzerland are considered attractive destinations for funds and property investment precisely because no one asks where the money comes from and laws punishing tax fraud do not exist (van Hulten, 2007), but does this make them as clean as their high rankings suggest? Are these the best practices to be copied? The subjective character of the indicators thus undermines their utility as tools for policy change.

What is more, two countries with similar rankings may suffer from different types of corruption; 'sometimes politicians are bent and judges are straight; civil servants may be honest executors of a corrupt politician's will; or the Minister may be honest, but officialdom crooked' (*The Economist*, 2006). The CPI does not distinguish between grand and petty corruption, although the first is presumably more threatening to the economy than the latter (Fjeldstad and Andvig, 2001: 29). As a result, the ranking does not give information about the kind of measures most suited for fighting corruption and may even lead to perverse outcomes instead of actually

improving governance. For instance, government involvement in private markets is commonly seen as a source of corruption (although this does not hold true for Scandinavian countries), but deregulation is not necessarily a solution. Deregulation may even result in 'privatized forms of corruption' (Lambsdorff, 1999: 2) that have the advantage of not being counted as corruption but are not connected to better governance. Another example of a perverse effect is that governments interested in improving their ranking for fear of losing foreign investments will be 'doing the document rather than doing the doing' (Ahmed, 2007) and, for instance, create 'a Commission for the Modernisation of the Civil Service' without dealing with the underlying structural causes of corruption (Urta, 2007).

Discussion

In this article we have shown that there are valid reasons to doubt the effectiveness of ranking and benchmarking in improving policy making in terms of effectiveness and accountability. All of them have to do with the political, power-based logic underneath the construction and application of these instruments. This section brings together our findings in the field of European gender equality policies and global anti-corruption policies.

First, best practices hamper rather than promote good performance, because their link to performance quality is equivocal. Past reputation tends to be rewarded without taking into account the context-specific character of any 'best' practice. Accordingly, best practices discourage states with other practices to improve their performance, as their practice is not recognised, while best practices enable those with the 'right' style to do window-dressing without incentivising performance improvement. This shift favours certain states and prejudices others. In the field of gender equality, thanks to its tradition of pilot projects and corresponding rhetoric, a state such as the Netherlands will be more likely to be considered a pioneer than a state such as Austria with its legalistic governance style and fewer catchy projects – although pilot projects not necessarily result in better *de facto* performance than legal instruments. In the field of anti-corruption, market-based economies have an advantage over economies with high state involvement. Yet, there is no straightforward connection between these elements and the quality of performance. The former will be considered the standard to be copied, not the latter, even if Scandinavian states with their strong state involvement in the market display lower levels of corruption than the US with a low level of state involvement. We have also argued that states with a good reputation are in a good position to influence the design of new indicators, ensuring their national practice to be optimally 'graded' by these indicators, and further strengthening their reputation. Similarly, states can be 'immunised' by a good reputation.

Second, the practice of ranking based on one-dimensional indices favours policies that are narrowed to quantifiable objectives, ignoring multiple linkages. The main criticism of legal standards was that legal implementation was too limited to realise complex policy goals such as gender equality or good governance. Ranking, however, does not remedy this problem, as it limits implementation to the quantifiable aspects of gender equality and good governance, and provides no information on the underlying causes or possible remedies of gender discrimination and corruption.

This limitation to what is measurable reproduces older patterns of power-based data collection, in which what fits in with the mainstream has the highest chance of being collected.

Third, ranking does not necessarily increase political accountability. At the global level, democratic control of standard-setting practices by international organisations is non-existent and neither are there possibilities for appeal. In the EU, negotiations on new legal standards to be approved by the national and European parliaments have been replaced by the production of seemingly depoliticised data. The latter is presented as a technical, a-political process in which experts determine the indices based on shared academic knowledge. However, it is a two-stage political selection process, first one of who will count as experts, and then one of negotiations between officials and experts in the selection of indices and the definition of targets, all without control by the national parliaments or the European Parliament and without possibility for appeal before the Court.

Of course, the fact that indices do not correctly assess policy performance is not new. It is also generally admitted that every index has its drawbacks. We, however, argue that this is often erroneously considered to have technical causes, such as a lack of data or expertise, requiring a technical solution, as if more data and expertise could produce an objective, absolute standard and as if the academic community and the involved experts (femocrats, TI) have no parochial interests in the production of indicators. It should be recognised that the choice of an index is always a political decision, therefore necessitating transparency as to who decides on the choice of indices and which political actors are involved, so they can be held accountable for these decisions.

Fourth, ranking and benchmarking imply a shift from absolute to relative and pioneer standards. Targets and standards are selected in terms of existing practices and performance is judged by peer review. This favours pioneers, while it obscures the goal to be attained by the laggards: is the policy objective to copy the number one or to attain the mean?

Fifth, ranking lacks an actor that enforces compliance for all parties equally and has an interest in compliance. There is no authoritative third party for the interpretation and enforcement of a standard, equal for all. When 'relative' monitoring replaces 'absolute' monitoring and compliance can only be enforced through naming and shaming, a mechanism is created that is based on the fallacy that states perceive themselves as a collective that is eager to learn and committed to finding the 'best' solution to any challenge it faces. The eagerness to learn is a necessary component of peer review. Assuming shared norms on harmony replaces the political arena – in which negotiations based on preferences and power lead to compromises – with a depoliticised arena in which national practices are harmoniously coordinated between equals. Instead, states may have a stronger interest in their joint reputation as 'gender-friendly states' or 'corruption-free Western states' than in blaming each other. The costs of sanctions are thus much lower than under the enforcement mechanism.

In sum, we contend that the politics of ranking have perverse effects on the link between performance and reputation. In spite of their aim to improve performance through the potential consequences of ranking for states' reputations, they still may hinder the improvement of the quality of performance and actually impede the

attainment of policy goals or the increase of the transparency and legitimacy of the political process.

Notes

¹ CPI scores range from 10 (extremely clean) to 0 (highly corrupt).

² Clarke (2000) and van Hulten (2007) also criticise the data collection: TI does not use the same sources every year; the size of the survey varies between the countries; and only English-language sources are used. Their criticisms further undermine the reliability of the index.

³ In spite of these shortcomings, the index is used by academics in cross-country empirical studies without questioning the quality of the data (Fjeldstad and Andvig, 2001: 64).

⁴ Former President Jacques Chirac and former Chancellor Helmut Kohl have been accused of having had a role in party financing corruption scandals ('Chirac questioned in scandal from his time as Paris mayor', *New York Times*, 20 July 2007; 'German sleaze: the story so far', *BBC News*, 17 February 2000).

References

- Ahmed, S. (2007) "'You end up doing the document rather than doing the doing": diversity, race equality and the politics of documentation', *Ethnic and Racial Studies*, 30: 590–609.
- Bacchi, C. (1999) *Women, policy and politics*, London: Sage Publications.
- Barysch, K., Tilford, S. and Whyte, P. (2008) *The Lisbon scorecard VIII: Is Europe ready for an economic storm?*, London: Centre for European Reform, www.cer.org.uk/publications_new/806.html
- Borrás, S. and Jacobsson, K. (2004) 'The open method of coordination and new governance patterns in the EU', *Journal of European Public Policy*, 11: 185–208.
- Börzel, T.A. (2003) 'Guarding the treaty: the compliance strategies of the European Commission', in T.A. Börzel and R.A. Cichowski (eds) *The state of the European Union, vol. 6*, Oxford: Oxford University Press.
- Caporaso, J.A. and Wittenbrinck, J. (2006) 'The new modes of governance and political authority in Europe', *Journal of European Public Policy*, 13: 471–80.
- Clarke, R. (2000) 'Transparent tables', *OECD Observer*, no 220, July, www.oecdobserver.org/news/fullstory.php/aid/296/Transparent_tables.html
- de la Porte, C. and Nanz, P. (2004) 'The OMC – a deliberative-democratic mode of governance? The cases of employment and pensions', *Journal of European Public Policy*, 11: 267–88.
- de Vries, M.S. (2005) 'Generations of interactive policy-making in the Netherlands', *International Review of Administrative Sciences*, 71: 577–91.
- Espeland, W.N. and Sauder, M. (2007) 'Rankings and reactivity: how public measures recreate social worlds', *American Journal of Sociology*, 113: 1–40.
- EC (European Commission) (2009) *Commission staff working document: Accompanying document to the report Equality between women and men – 2010 (COM(2009) 694 final)*, Brussels: EC.

- Fjeldstad, O. and Andvig, J.C. (2001) *Corruption: A review of contemporary research*, CMI Report No 2001:7, Bergen: Chr. Michelsen Institute, www.cmi.no/publications/
- Haque, M.S. (2007) 'Revisiting the New Public Management', *Public Administration Review*, 67: 179–82.
- Héritier, A. (2003) 'New modes of governance in Europe: increasing political capacity and policy effectiveness?', in T.A. Börzel and R.A. Cichowski (eds) *The state of the European Union*, vol. 6, Oxford: Oxford University Press.
- Joachim, J., Reinalda, B. and Verbeek, B. (2008) 'International organizations and implementation: pieces of the puzzle', in J. Joachim, B. Reinalda and B. Verbeek (eds) *International organizations and implementation: Enforcers, managers, authorities?*, London: Routledge.
- Lambsdorff, J.G. (1999) *Corruption in empirical research: A review*, Passau, Germany: ICGG, www.icgg.org/downloads/contribution05_lambsdorff.pdf.
- Lombardo, E. (2003) 'EU gender policy trapped in the "Wollstonecraft Dilemma"', *European Journal of Women's Studies*, 10: 159–80.
- Lombardo, E., Meier, P. and Verloo, M. (2009) *The discursive politics of gender equality: Stretching, bending and policymaking*, London and New York, NY: Routledge.
- Luxembourg Presidency (2005) *Report of the Luxembourg Presidency on the progress made by the enlarged EU following the Beijing Platform for Action (Beijing +10)*, Prepared by A.M. Theisen, N. Spoden, M. Verloo and S. Walby, Luxembourg: Luxembourg Presidency.
- Majone, G. (1997) 'From the positive to the regulatory state: causes and consequences of changes in the mode of governance', *Journal of Public Policy*, 17: 139–67.
- Merton, R.K. (1968) 'The Matthew effect in science', *Science*, 159 (3810): 56–63.
- Mósesdóttir, L. (2006) Final report: Research results of the project: From welfare to knowfare. A European approach to employment and gender mainstreaming in the knowledge based society, Sími, Iceland: Bifröst University, www.bifrost.is/wellknow/Files/Skra_0014164.pdf
- Mosher, J.S. and Trubek, D.M. (2003) 'EU social policy and the European Employment Strategy', *Journal of Common Market Studies*, 41: 63–88.
- Plantenga, J., Figueiredo, H., Remery, C. and Smith, M. (2003) *Towards an EU gender equality index: Feasibility study commissioned by and presented to the European Commission*, Manchester: Manchester Business School, www.mbs.ac.uk/research/europeanemployment/projects/gendersocial/documents/EU%20index%20final%20report%20December%202003.pdf
- Pfister, T. (2008) 'Mainstreamed away? Assessing the gender equality dimension of the European Employment Strategy', *Policy & Politics*, 36 (4): 521–38.
- Roggeband, C. and Verloo, M. (2006) 'Evaluating gender impact assessment in the Netherlands (1994–2004): a political process approach', *Policy & Politics*, 34 (4): 615–32.
- Rönblom, M. (2004) 'The best place for women to be? A study of the construction of gender equality and gender mainstreaming in regional policies in Sweden and the EU', Paper presented at the 2nd Pan-European Conference on EU Politics, 24–25 June, Bologna, Italy.

- Rubery, J. (2005) 'Gender mainstreaming and the open method of coordination: is the open method too open for gender equality policy?', in J. Zeitlin and P. Pochet (eds) *The open method of coordination in action: The European Employment and Social Inclusion Strategies*, Brussels: PIE-Peter Lang.
- Rubery, J. and Fagan, C. (2000) *Gender impact assessment and European Employment Strategy*, Manchester: EWER/UMIST.
- The Economist* (2006) 'Strains of sleaze: how trustworthy is the experts' verdict on governments' honesty?', 9 November.
- Transparency International (2006) *BPI 2006: Leading exporters undermine development with dirty business overseas*, Brussels: Transparency International, www.transparency.org/news_room/in_focus/bpi_2006
- Transparency International (2007) *TI Corruption Perceptions Index*, Brussels: Transparency International, www.transparency.org/policy_research/surveys_indices/cpi/
- Urta, F.J. (2007) 'Assessing corruption: an analytical review of corruption measurement and its problems: perception, error and utility', <http://unpan1.un.org/intradoc/groups/public/documents/APCITY/UNPAN028792.pdf>
- van der Vleuten, A. (2005) 'Pincers and prestige: explaining the implementation of EU gender equality legislation', *Comparative European Politics*, 3: 464–88.
- van der Vleuten, A. (2007) *The price of gender equality: Member states and governance in the European Union*, Ashgate: Aldershot.
- van Hulten, M. (2007) *Ten years of Corruption (Perceptions) Indices: Report*, Enschede: Corruptie, www.corruptie.org/
- Verloo, M. (2005) 'Displacement and empowerment: reflections on the Council of Europe approach to gender mainstreaming and gender equality', *Social Politics*, 12 (3): 344–66.
- Verloo, M. (ed) (2007) *Multiple meanings of gender equality: A critical frame analysis of gender policies in Europe*, Budapest: CEU Press.
- Walby, S. (2005) 'Gender mainstreaming: productive tensions in theory and practice', *Social Politics*, 12: 321–43.

Anna van der Vleuten (a.vandervleuten@fm.ru.nl), Institute for Management Research, Radboud University Nijmegen, The Netherlands

Mieke Verloo, Institute for Management Research, Radboud University Nijmegen, The Netherlands and Institute for Human Sciences, Vienna, Austria