The discursive logic of ranking and benchmarking

Understanding gender equality measures in the European Union

Mieke Verloo and Anna van der Vleuten

Introduction

Strategies to promote gender equality are developing continuously. While the main strategies remain equality in legislation, positive or affirmative action and gender mainstreaming, the last decade has seen a proliferation of instruments to measure and monitor progress towards a gender-equal society. Resonating well with ‘new public management’, the introduction of indicators and more complex indices results in country rankings that are seen as indicative of their achievement of gender equality. The recent increase in such rankings makes them more important for the assessment of countries’ performances in this field.1 A common feature of ranking and benchmarking is the underlying idea of the concern for reputation as a motor for implementation.

At first sight, gender rankings based on indicators seem to be a prime example of ‘fixing’ the meaning of gender equality. As outlined in the Introduction, such fixing can have positive and negative consequences. While it can enable the evaluation of progress, and create a chance to take measures if there is a lack of progress, it is not unlikely that the indicators represent only a few dimensions of gender equality. These instruments possibly show evidence of shrinking the meaning of gender equality, or stretching or bending part of its meaning.

This chapter will address the discursive politics of ranking and benchmarking, because we consider the discursive dimension of these phenomena to be under-valued. Ranking appears to be a technical undertaking based on expert knowledge and depoliticized facts and figures, and is often discussed as such. Yet, power is unavoidably as central to it as to other policy instruments. This has led us to investigate the origins and consequences of what we consider shifts in ‘the politics of reputation’ as a result of the increased importance of new policy methods such as ranking. We examine the power-based mechanisms that operate between each pair of the following three concepts: reputation, ranking and performance. How does reputation relate to ranking? A good reputation seems to influence the way in which rankings are constructed as much as rankings are constitutive of a good reputation. How does performance relate to reputation, and to what extent do we see disparities between them? Does ranking improve or hamper good performance? In this chapter, we try to conceptualize the consequences of the ‘new’ politics of
ranking in order to contribute to a political understanding of their impact on the meaning of gender equality and on the quality of policy implementation.

Our subject is gender equality policies in the European Union (EU). In the 1990s, this field has seen a shift in policy and implementation style from legal instruments aiming at harmonization of national policies, enforced by the European Commission and the European Court of Justice (ECJ), to policy instruments such as the co-ordination of national policies (Open Method of Coordination, European Employment Strategy), benchmarking and scoreboards (Héritier 2003). Here, the European level reflects broader changes within countries and in the private sector (see Benschop and Brouns 2006 on the Netherlands’ Glass Ceiling Index, which ranks companies; or Grosser and Moon 2005 on corporate responsibility). Although the emergence of ranking as a policy measure can be found in many domains at the European level, its effects on reputation and performance can be assumed to be stronger if the policy domain is less ‘technical’ and more ‘ideological’, i.e. linked to power relations and policy paradigms. For that reason, gender equality policies constitute the domain par excellence where the political aspects of these mechanisms – if any – should be most visible.

In the next section, we explore the mutual influence between ranking, reputation and performance. Subsequently, we will trace the development of ranking in the field of gender equality in Europe with special attention to the role of power and some particular effects of the politics of ranking. In closing, we will discuss how this new type of tool affects the meaning of gender equality, how stretching and bending are as much active in these tools as in any other type of strategy, and how existing power structures influence the construction and use of indicators, benchmarking and ranking.

Reputation, ranking and performance

States value their reputation. Rönnblom (2004: 2) notes how Sweden proudly presents itself in policy documents as the ‘most gender equal country in the world’. Reputation is part of the power position of a state with regard to domestic society and to other states (van der Vleuten 2005). A good reputation legitimizes the course a government has chosen and may also serve to legitimate further inaction as a good reputation, once acquired, lasts a while. It is therefore an asset in international negotiations as well. If others are convinced of the qualities of a state in the past, this state is able to convince them more easily of the quality of its present ideas. Conversely, ‘being perceived as a cheater and free-rider undermines the bargaining power of a member state’ (Börzel 2003: 203).

Reputation is based not only on actual past performance, but also on the way in which this performance is represented. Although this rhetorical dimension makes reputation malleable to a certain extent, reputation has to remain rooted in facts to avoid a sudden loss of credibility when the facts are checked. In effect, rhetorical action (defined as the strategic use of arguments) follows some rules, described by Schimmelfennig (2001) as requirements of consistency, credibility and a need to appear convincing, with a resulting preference for obscuring or hiding
inconvenient facts or norms. Schimmelfennig also points to an interesting unintended consequence of rhetorical action. Even if a state uses rhetoric to enhance its reputation in order to gain more leeway in negotiations at the European level, its *marge de manoeuvre* may be involuntarily limited by the ensuing rhetorical entrapment (Schimmelfennig 2001).

The concern for reputation plays a role in explaining state compliance with ‘hard’ EU law. If a state prefers not to implement a policy, it will prefer this non-compliance to remain unnoticed, so that its reputation as a reliable member state does not suffer. As the EU monitoring system increases the probability that non-compliance is discovered, the threat of ‘being named and shamed’ is supposed to influence the decision of a state to implement EU measures (Tallberg 2002).

The ‘old’ politics of reputation consisted not only of monitoring but also – in the case of hard law directives – of enforcement measures. These enforcement measures enable the European Commission to investigate whether state rhetoric corresponds with the facts, and to unveil instances of non-compliance by taking a state to the ECJ. With each stage in the procedure, the publicity increases, as well as the ‘reputational costs’ (Börzel 2003). Member states prefer not to have infringement proceedings in front of the ECJ, as the European Commission usually wins the case and ECJ rulings receive a lot of publicity, ‘further raising the social costs of non-compliance’ (Tallberg 2002) and damaging their reputation. The involvement of the ECJ implies that an ‘absolute’ standard is applied, that is the standard as laid down in the specific European policy. As a third party, the ECJ interprets this standard, making it applicable to all member states, businesses and citizens in the same way. This enforcement mechanism enables the denunciation of disparities between reputation and performance. France, for instance, cherished its reputation as a social role model, defending gender equality from the very beginnings of European integration. The European Commission, however, took France to court more than once because it did not fully comply with the gender equality directives. When in 2000, just when it occupied the EU presidency, France risked being condemned to a daily fine of €142,425 if it did not implement the equal treatment directive, the French government succeeded in convincing the French parliament of the necessity to act immediately by pointing not at the financial consequences, but at the consequences for its reputation. To be the first member state ever to be condemned to a penalty payment over social policy was unacceptable for France as the would-be champion of a ‘social Europe’ (van der Vleuten 2007: 122).

The ‘old’ politics of reputation, linked to legal performance, is thus based on ‘absolute’ standards authoritatively interpreted and enforced by a third party (the European Commission and the ECJ). Such absolute standards are a prime example of fixing the meaning of gender equality, necessarily involving only certain parts of its meaning. Equal rights in the labour market and equal pay are good examples of this, fixing ‘equality’ in a more narrow sense to be part of gender equality, while completely ignoring the division of care and household labour, or issues of reproductive rights. It therefore seems that absolute standards involve shrinking as well as fixing.
In spite of the strong monitoring and enforcement system in the EU, differences persist when it comes to the legal performance of member states. Looking at the implementation of EU gender equality directives, some states are considered pioneers, while others are labelled laggards or poor performers (Liebert and Sift 2003; Falkner et al. 2005). This kind of labelling focuses solely on legal equality between women and men, a strategy that has been criticized for not always being sufficient for the attainment of de facto gender equality (Lombardo 2003). For instance, although women have for decades been legally entitled to receive the same pay for the same work as men, the gender pay gap has not disappeared. Accordingly, we agree that the standard of ‘legal implementation’ is not an appropriate standard if one wants to measure actual performance (the attainment of policy objectives), but at least it is an ‘absolute’ standard that is applied to all states in the same way, making it possible to judge them accordingly.

Conversely, the ‘new’ politics of reputation linked to ranking is based on relative standards of performance and excludes the possibility of judging and sanctioning low performance. While reputation is an old concept in, among others, theories of state behaviour, analyses of political leadership and management theories, ranking is a relatively new one. Since the 1980s, new public management and other similar trends in professionalizing bureaucracies have given birth to a panoply of indicators, procedures, measures and other innovations meant to assess the performance of public administration systems and to improve the performance and political accountability of governments (de Vries 2005). Intended to be instruments for improving the quality and transparency of policies, they have given new meaning to the old politics of reputation by changing both the basis on which reputation is built and the mechanisms by which reputation is judged.

**Ranking**

The relation between reputation and ranking varies, as ranking refers to different practices. It includes the establishment of indicators, which disaggregate policy goals in quantitative objectives. Based on their scores on such an indicator, states may then be ranked and classified as pioneers or laggards. Under conditions of measuring and ranking performance, reputation is no longer based mainly on shared understandings between state and society, and among states, but has to be ‘proven’ in technocratic processes. The development of technocratic processes to measure performance implies many decisions as to which aspects of a phenomenon will and can be measured. In policy practice, this has substantial consequences, as different states will be at an advantage or disadvantage depending on the chosen ways of measuring. It then also matters which procedures are chosen to select indicators and who has the power to decide what data have to be produced to fill them.

Another method of ranking consists of selecting best practices, pilot projects and benchmarks. Best or good practices are assumed to offer the best solution to a problem and are then presented as the model to be copied. This method does not result in an absolute standard that all states have to apply in order to avoid
coercive measures in case they do not conform. Even shaming will be of limited use, as a good reputation does not necessarily presuppose the proposition of best practices, and the development of different practices does not necessarily lead to failing to solve the problem for which the ‘best practice’ was recommended. Nonetheless, there might be an interesting connection between this aspect of ranking and reputation that goes in the opposite direction: to launch a new strategy or tool, an institutional organization may decide to collect best practices, even if examples of the new strategy or tools are barely present in actual practice. The result to be expected if this happens is that practices proposed by states with a good reputation in the field will have more chance of being adopted than practices proposed by states that are perceived as laggards. Therefore the standard of ‘best practice’ models is neither a generalized absolute standard nor a relative one, but the specific standard of a state considered to be a pioneer (‘pioneer standard’), thus further confirming the reputation of the pioneer.

We argue that the current politics of ranking, at odds with its presented ambitions of improving performance and policy learning (Trubek and Trubek 2005), works against the quality and even the transparency of policymaking, because its objectives are unidimensionally quantified and its concepts and indicators are chosen on the basis of availability of data, without any transparent democratic debate. The consequences are that concepts are stretched and bent to fit the available data. As a highly contested problem, gender equality is easily subject to such processes. As a result, the assessment of performance is of poor quality; it is unclear what kind of performance is actually measured or how performance could be improved. As the chosen indicators are not only informative about progress on gender equality, but also tell a story about the power relations among member states and between member states and the EU, this story needs to be unravelled to show the full effects of the politics of ranking.

**Practices in measuring and benchmarking gender equality**

Gender equality is a domain in which the EU has strongly influenced national policies since the 1970s: it has developed directives that had to be transposed in national legislation; it has promoted the establishment of transnational networks; and it has financed projects on equal pay, equality and employment, and sexual violence (van der Vleuten 2007). Since the 1990s, the range of mechanisms to promote gender equality within European countries and at the EU level has been expanded from equal treatment in legislation – focused on providing equal access to the labour market – to a strategy of targeted gender equality policies, which often take the form of temporary projects for specific groups of women, and gender mainstreaming (Verloo 2001). There is also an increasing use of gender indicators and benchmarks (Luxembourg Presidency 2005). At the level of the EU, their introduction was part of the development of the Open Method of Coordination (OMC) and the European Employment Strategy (EES) and, in the field of gender equality policies, these European developments received a strong impulse through the Beijing review process, especially the Beijing +5 and the Beijing +10
reviews in 2000 and 2005 (Rubery 2003). Last but not least, the Organization for Economic Co-operation and Development (OECD) recently presented a new index (Gender, Institutions and Development, GID) that combines attention to both economic development and institutions. This index is relevant for European countries too, and includes some innovative features.

**Gender equality in the OMC and EES**

The OMC and EES processes have given member states new opportunities to represent their performance in the field of gender equality, and hence also their reputation. Gender equality has become embedded as a key component of the EES (European Council 1997, 2000). In its ambition ‘to become the most competitive and dynamic knowledge-based economy in the world, capable of sustainable economic growth with more and better jobs and greater social cohesion’, the European Council (2000: 2) has integrated gender equality in the guidelines by which member state progress on economic policy is to be assessed at the EU level. Interestingly, an important impulse to the inclusion of gender equality in the EES has been given by a country that does not exactly have a strong reputation in this field. According to Rubery:

> the Austrian ministers for women and for labour in post during the 1998 Austrian presidency, both of them women ministers, had their sights set on promoting equal opportunities further within the EES. To that end they organised an informal joint council of ministers for labour and gender equality, commissioned an extensive report and used this momentum to push for the successful inclusion of the commitment to gender mainstreaming in the EES guidelines at the December summit.

(Rubery 2003: 3)

For the assessment of progress in realizing the Lisbon ambitions, a new method of policy development, the open method of coordination (OMC), has been put in place. The OMC is a semi-voluntary form of co-ordination (Scott and Trubek 2002). It contrasts with the ‘old’ Community method in that it relies on a mix of ‘soft law mechanisms’ such as peer review, best practices and monitoring. Targets for member states are set at the meetings of the Spring European Council (using structural indicators, overseen by Eurostat). The Council organizes the annual monitoring of progress towards these targets through peer review, and the European Commission issues recommendations to individual member states. While the European Council sets the targets, each member state can develop its own strategy to reach the targets, paying attention to its national specificities. To promote mutual learning, a process of identification and exchange of best practices is organized between the member states, and guidelines based on these best practices are then incorporated in the National Action Plans. The implementation of the guidelines is monitored in the Joint Employment Report (JER), and leads to a renewed identification of best practices and to a new start of the OMC cycle (Biagi 2000). As part of their
mutual learning strategy, the OMC and EES processes aim to foster mutual learning and the generation and diffusion of new ideas and practices across Europe. This postulates the member states as a collective that is eager to learn and committed to finding the ‘best’ solutions to any of the challenges ahead. While this postulation can be questioned on the basis of past performance of states in implementing gender equality directives (van der Vleuten 2007), and there are fierce debates as to whether or not the OMC is delivering on its promises (Héritier 2003; Trubek and Trubek 2005), it is possible to draw some basic conclusions from the decade-long existence of this method for the field of gender equality.

For gender equality, the most important commitments made by the European Council are those concerning the narrowing of the gender gaps in employment (by increasing women’s employment rate to 57% by 2005 and to 60% by 2010) and unemployment (European Council 2000), as well as the provision of childcare for at least 90% of children between 3 years and school age and at least 33% of children under 3 (European Council 2002), and the narrowing of the gender and pay gap (European Council 2003). A basic result of the OMC and EES processes is that there is an ongoing production of rankings of the member states on employment and employment-related issues. The targets, however, only have ‘a general orientation function’ (Héritier 2003: 118), and member states are pressured but not compelled to meet them. The peer review as well as the evaluation and recommendations by the European Commission constitute a monitoring mechanism without sanctions, other than ‘naming and shaming’ through publication of the performance schedules. For an underperforming state, the reputational costs of the OMC and EES processes therefore remain limited compared with the costs occasioned by the enforcement system linked to legal instruments.

Another result is the production of best practices in the field. Member states are ‘encouraged to benchmark their performance against the best performer in the Union’ (Héritier 2003: 117). This implies that one of the national solutions is considered the best one, without questioning all practices and investigating which practice suits which national context best. National institutions are not called into question in the same way in which legal instruments could oblige a state to revise its central concepts or practices: the way in which, for instance, the equality directives required the Dutch and British governments to fundamentally revise their breadwinner-centred systems of pay and social security in spite of their arguments that this system served women’s needs well enough (van der Vleuten 2007).

In addition to the information given by member states, the European Commission uses expertise from networks of gender experts to assist in its assessment of member state progress in this area. They are another source of best practice ‘labelling’, creating ‘pioneers’ in the process of doing so. In their review of all the performed pan-European gender impact assessments (GIA), Rubery and Fagan (2000) position the Netherlands as an early starter on GIA, together with the Nordic countries except Denmark. They then incorporate the Dutch conceptual framework for GIA as an academic model for what they recommend as a required further conceptual and practical elaboration of the GIA methodology (Rubery and Fagan 2000: 5). In practice, however, the Dutch GIA methodology has proven to be difficult to
export because it is tailored to the Dutch context (Roggeband and Verloo 2006). This shows that best practices have no value if they are not contextualized.

**The Beijing review process**

Another important impulse to the development of measuring and benchmarking gender equality was given by the follow-up to the 1995 Beijing United Nations (UN) conference. Under the leadership of the UN, the Beijing Conference in 1995 resulted in a Platform for Action to bring about gender equality in the world, which has been adopted by all EU countries. At two points in time, in 2000 and in 2005, the UN took further action to monitor the progress made. On each occasion, governments were asked to report on their actions to implement the Platform for Action in the twelve critical areas of concern. In 2000, the European Parliament adopted a Resolution acknowledging:

> that various aspects of women’s life in the EU have improved since 1995, whilst deploring the lack of agreed indicators, gender-segregated data, benchmarks and a clear timetable in the Platform for Action, this being a major obstacle to proper evaluation of progress made over the last five years [and urging] the participants at the New York Conference on Beijing +5 to establish indicators, benchmarks and a precise time schedule.

(European Parliament 2001: 260)

This resolution clearly shows the importance of the Beijing review process for the development of pressure towards measurable output of gender equality policies at the European level.

In recent years, the European Council has presented itself as playing a leading role in the development of indicators to operationalize the specific gender equality objectives (Luxembourg Presidency 2005). In 1998, the Council committed the EU to the development of a simple suite of indicators to monitor progress on all twelve critical areas of the UN Platform for Action. Although the set of indicators is still not complete, there have since been a number of proposals by a series of EU Presidencies: Finland 1999, women in power and decision-making; France 2000, reconciliation of family and working life; Belgium 2001, gender pay gap; Spain, Denmark, Ireland, Greece, Netherlands, 2002–4, violence against women. In addition, as described earlier, some indicators related to employment have been agreed on by Councils as part of the EES. The EU now has indicators in five of the twelve critical areas: women and the economy (gender gaps in employment, unemployment and pay; provision of care for children and other dependents); women in power and decision-making (percentage of women among elected positions, ministerial positions, executive boards of top companies and boards of central banks); women and poverty (percentage of women among those “at-risk-of-poverty”); women and education (educational attainment, life-long learning, science and technology graduates); and violence against women (domestic violence: number of female victims, types of victim support, measures to end violence, sexual harassment at work).
In the period after 1995, progress reports on these issues have been prepared by the European Commission for the annual reviews of different critical areas of the Platform for Action by the UN Commission on the Status of Women (CSW).

**Recent innovation: the OECD GID index and database**

The new OECD index on Gender, Institutions and Development (GID) includes variables on the economic development of countries, the economic role of women, the access of women and men to resources in health and education, as well as a set of variables on social institutions that are seen as crucial for gender equality (Jütting *et al.* 2006). The variables on social institutions are about family code (early marriage, repudiation practices, parental authority and inheritance practices), physical integrity (female genital mutilation, legislation on violence against women), civil liberties (percentage of women in parliament and among ministers, women’s freedom to leave the house, requirement to wear the veil in public) and ownership rights (access to bank loans, landownership and right to own non-land property). Most of the social institutions included in GID deal with legal arrangements, except for the indices on early marriage and female genital mutilation. This bias has been criticized because it suggests that mainly formal institutions matter in restricting the position of women (Van Staveren 2008). One could also question the choices made, as the chosen indicators seem biased against Islamic countries (‘requirement to wear the veil in public’ as an indicator of civil liberties), where they could just as well have chosen an indicator of social safety in public spaces that would have shown more negative results for Western countries. Including an indicator on repudiation practices rather than on the possibilities and conditions for divorce seems biased in a similar way.

Countries are positioned very differently on the GID rankings than on the UN’s GDI and gender empowerment measure (GEM) rankings. Most notably, the United Kingdom, Ireland and Germany are in second or shared third position on the GID index, as opposed to a sixteenth and seventeenth position for the UK and Ireland on GEM and a twenty-first position for Germany on GDI. Iceland and Finland, on the contrary, although high-ranking on the UN indices, fall down to a shared thirty-sixth position on the GID. In the top ten of the GID, Korea is the surprise, ranking eighth. It is beyond the scope of this chapter to analyse which indicators of the index privilege which country.

To some extent, the OECD anticipates and facilitates debate over its index, by designing a new portal, Wikigender, which they hope will be instrumental in reaching out to the public and fostering a bottom-up dialogue on the importance of gender equality. The website aims to provide:

- an open forum to share and exchange information on the situation of women and [encourage] a frank discussion on the elements that prevent women’s social and economic empowerment. It welcomes an active participation of users who can contribute to the content of the website by posting comments, editing articles or creating new entries into this knowledge database.

(Drechsler and Jütting 2008: 8)
As the portal has only opened recently, it is too early to assess whether these ambitions will materialize.

The discursive politics of performance

Having traced the development of new policy methods in the field of gender equality, what are the assets and drawbacks in the discursive politics of performance? The usual defence of the necessity of measuring gender equality and of developing gender equality indicators and the data to support them – whether done by policymakers or by academics – is that they are necessary to know whether gender equality policies are effective (Plantenga et al. 2003; Luxembourg Presidency 2005). It is no surprise then that there are complaints about the lack of indicators across Europe and about the lack of data to support the indicators (Luxembourg Presidency 2005).

This chapter has set out to contend that the increasing use of gender indicators and benchmarks resulting from the EES and Beijing review processes presents a number of serious problems, even while it no doubt also contributes to a higher visibility of gender equality and potentially supports progress towards gender equality. These problems result from power processes in the construction or use of indicators, and lead to discursive dynamics. It is to these dynamics that the next section turns, discussing the shift to relative standards, the selection bias in the development of targets, benchmarks and indicators and the technocratization of performance measures.

The shift to relative standards

In spite of Europeanization and globalization, differences between states persist, so the annual reports show. The presentation of states’ performances on the chosen indicators seems to ‘prove’ which states are doing well and which states are underperforming. However, the ranking that results from these indicators is clearly problematic. The EES’ absolute targets are not de facto targets for everyone, as some member states are outperformers on these targets from the start. The indicators are used for relative ranking of member states, and the ‘EU average’ is presented as an important benchmark for member states. Mosesdottir argues that:

> the emphasis made by the EES of ranking countries according to their performance on narrowly defined targets like, for example, the female employment rate, runs the risk of making those countries that out-perform the others the unachievable standard for member countries falling behind. Economic, political and social contexts cannot be copied overnight. Moreover, those nations performing well tend to focus more on their relative position than on their distance from the EU’s employment and gender objectives.

(Mosesdottir et al. 2006: 53)

The shift from legal instruments to scoreboards and benchmarks also favours states with a policy-oriented style. Thanks to its pilot projects and corresponding rhetoric, a state such as the Netherlands will more easily be considered a pioneer
than a state such as Austria, with its legalistic governance style and less catchy projects, even though pilot projects do not necessarily result in better de facto performance than legal instruments. If analysed more closely, pioneers may therefore disappoint and laggards surprise; as described above, it was not gender champion Sweden but laggard Austria that took the initiative to include gender equality in the EES in 1998 to begin with.

The shift from ‘absolute’ to ‘relative’ standards means that there is no authoritative third party for the interpretation and enforcement of a standard that is equal for all; in its place, we find a power-driven process in which targets and standards are chosen as a function of existing national practices. ‘Relative’ monitoring (through peer review) replaces ‘absolute’ monitoring. There is no sanctioning mechanism apart from ‘naming and shaming’. Reputational costs therefore remain more limited. In addition, the third parties (the European Commission and ECJ) have an interest in assuring compliance with EU (gender equality) legislation in order to keep up their reputation as guardians of the Treaties, whereas member states in a peer review process have an interest in keeping up their ‘joint’ reputation and showing their commitment to gender equality, not in blaming each other. The politics of ranking therefore reduces the reputational costs of poor performance compared with those of the old politics of reputation. Politics of ranking thus allows pioneers to do window-dressing instead of assuring real implementation, whereas some so-called laggards can actually be performing really well if we focus on their de facto performance. Ranking distorts the value of performance and is therefore problematic as an instrument for the improvement of performance, as both pioneers and laggards lack an incentive to do better.

What gets stretched in these dynamics is the meaning of quality, and changing the meaning of quality from an absolute to a relative understanding also changes the meaning of gender equality to something that is achieved if other countries do worse. This again means shrinking gender equality to less than what an absolute standard would amount to, which is especially problematic for the European countries who do very well in most rankings.

Nevertheless, relative standards can also offer new discursive opportunities. Recently, a very low ranking of Cyprus by the US State Department was used by a feminist non-governmental organization (NGO) to express its alarm and concern towards the Cyprus government in relation to combating trafficking in women for the purpose of sexual exploitation (MIGS 2008). In using the discursive opportunities that rankings offer, NGOs can ‘pinch’ their governments and successfully pressure them to action (van der Vleuten 2005).

Selection bias in the development of targets, benchmarks and indicators

After the presentation of the OECD GID database, some striking differences in ranking positions across indices have already been presented and discussed. A feasibility study for an EU gender equality index commissioned by the European Commission shows in more detail how various indices can produce very different rankings (Plantenga et al. 2003). One index showing the equal opportunity
indicators would put the usual laggards Portugal and Belgium ahead of gender pioneers Denmark and Sweden, and would even put the often high-ranking Finland at position 9 (EU 15). The other index, based on a different set of equal opportunity indicators, 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). Their comparison also shows that a different way of measuring the gender gap in participation, using a headcount basis instead of a full-time equivalent, would lower the position of the Netherlands and the United Kingdom in the overall ranking (Plantenga et al. 2003: 47). The point here is not to argue about the respective quality of indices and indicators, although it is obvious that such questioning is useful and necessary, but to accentuate the fact that the choice of one or the other has substantial implications for the ranking and, consequently, for the reputation of states.

This points to the problem of the indicators’ implicit and explicit selection. The problem of selection bias is connected to the different ways in which states frame gender inequality as a policy problem and define gender equality as a goal. Consequently, some domains of gender equality that are close to mainstream fields of policymaking are emphasized over others in the construction of indicators. One can observe that there are more indicators about the public sector than about the private domain (EES), that inclusion will be more readily measured than structural progress and, in view of the absence of intersectionality in gender equality policies altogether (see Lombardo and Verloo in this volume), one can expect indicators to be based on the unitary category of women. Overall, it seems that indices rarely include indicators on sexuality, reproductive rights and relationships, with the exception of the GID index. Moreover, the accent on formal equality or on formal institutions can hide inequality in practice. Additionally, we can observe an absence of complex models integrating the indicators in the different fields, which obscures the complex interrelationships between gender inequalities in citizenship, knowledge, labour and intimacy.

A further difficulty related to benchmarks, targets and indicators is that they are inherently normative, and that most actors are well aware of this in their actions, even if they play it down for the sake of strategy. This triggers ‘reactivity’, the phenomenon of actors strategically playing into indicators to raise their scores, regardless of whether they actually perform well (Ahmed 2007; Espeland and Sauder 2007). This phenomenon has yet to receive adequate attention. Academic debates on the content of indices are rare, as are crossovers between academia and policymaking institutions (but see Plantenga et al. 2003; Walby 2005). In academia, large-scale comparisons would be needed for such assessments, and there is a lack of funding for such research. Democratic debates and academic studies on the political dynamics of ranking are even more scarce (with the notable exceptions of Ahmed 2007; Espeland and Sauder 2007). Experiences with evaluation or monitoring independent from state actors are also still exceptional.

The discursive dynamic not only influences the rankings that result from it. It also influences the politics of reputation. In 1997, a Protection Against Violence Bill entered into force in Austria. This is a law on domestic violence that is based
on an innovative model of conceptualizing state intervention in this field (Logar 2005). Six years later, after several other countries had adopted (parts of) what has come to be called the Austrian model, the Netherlands – an alleged pioneer – introduced a similar law after long studies and extensive debates. This example shows that what is seen as a laggard country can be the origin of much needed innovation. However, it has not changed Austria’s reputation as a laggard in gender equality policies. This could be connected to the fact that there are no targets, benchmarks or indicators on gender violence in the EES. After all, what is not measured does not count.

Similarly, other inequalities intersecting gender also do not count at this moment, and the framing of gender in indicators is often based on the unitary category of women. Attention to intersectionality in current indicators would actually challenge the Scandinavian countries’ great reputation for their woman-friendly policies and for their egalitarianism, as becomes clear in a special issue on gender and multiculturalism (Williams 2008) that severely criticizes these countries for their reinforcement of the otherness of non-Scandinavian cultures and their stereotypical views of minority women. Currently, however, in the absence of intersectional indicators, the reputation of the Scandinavian countries goes unchallenged.

**Understanding quality of performance as a technical matter**

Problems with indicators are often understood as a matter of lack of expertise in a technical sense (Luxembourg Presidency 2005). Some of them are seen to be connected to a need for adequate and comparable data. This lack of data is not only a practical or technical problem, but also a political one. Serious problems can arise when data need to be collected to ‘fill’ the indicators; it does indeed take several years before substantial indicators can be available and, although comprehensive monitoring is possible, this is mostly still subject to the decision-making power of the different actors involved. The lack of gender mainstreaming at the level of official statistical institutions is a barrier, as well as the confusion about who exactly has the power to start implementation of agreed-upon indicators in the EU: the Parliament, the Council, the European Commission, Eurostat or the member states.

Rubery (2003), while also representing the problem to be of a rather technical nature, gives two examples that show the political character of the use of benchmarks in the Joint Employment Reports (which provide the official review of national progress made towards achieving stated objectives). In one case, she notes, Greece was commended for closing the gender employment gap, but no reference was made to the fact that this was only achieved through a fall in the male employment rate. Similarly, changes in the segregation index were used to indicate progress in desegregating labour markets, without these changes being decomposed into effects related to structure change and labour force change or to changes in the gender shares within occupational categories. She rightly concludes that there is evidence of the misinterpretation of misleading indicators, but does not analyse this in terms of political processes and power. In this, she downplays these political processes and presents them as so-called technical ones.
As a result, seemingly depoliticized facts and figures, as translated in a position on the scoreboard, ‘prove’ the quality of performance, and negotiations on the quality of performance between the European Commission, Council and Parliament, or between governments, the European Commission and the ECJ, are becoming rarer.

We have also noticed that states with good reputations are able to strongly influence the design of further indicators in their areas of ‘excellence’. By being asked to send national experts to contribute to indicator design, they at least ensure that their national practice is optimally ‘known’, further strengthening their reputation. Hence, the choice of indicators is always a political decision, which therefore necessitates transparency as to who decides and to the involvement of political actors who can be held accountable for these decisions. Ranking is not merely a technical process but essentially a political process, which should be transparent and subject to democratic control. It would be interesting to reconstruct the decision-making process on the existing EU indicators to see to what extent not only the availability of data (a partly technical matter) but also the preferences of states have played a role in the choices that were made.

**Conclusion: The perverse politics of performance and ranking?**

Here, we set out to discuss the consequences of shifts in policy instruments for the ‘politics of reputation’. We have explored the power-based configuration of performance, ranking and reputation that was created by these shifts, as well as their perverse effects on performance. In spite of the aim of improving performance through ranking’s potential consequences for reputations, the politics of ranking may therefore hinder the improvement in the quality of performance in terms of attaining policy goals and increasing the transparency and legitimacy of the political process. We contend that the politics of ranking has perverse effects on the link between performance and reputation.

We have found specific stretching and bending processes in the construction and the use of benchmarks, indicators and rankings. The meaning of quality is stretched in a way that makes it relative. Most of all, the representation of these instruments as purely technical hinders democratic deliberation and struggle. It would be preferable to think of indicators, rankings and benchmarking as moments in processes of struggle (for the antagonistically inclined) or of learning (for the deliberatively inclined). This would highlight the discursive opportunities that these instruments can offer, and make the indices work as instruments for reflexivity as well. Indicators need to be continuously improved and developed in transparent democratic procedures, and independent actors are needed to interpret the results of the benchmarks, targets and indicators. More academic research parallel to, and critical of, these indices would also be beneficial. The OECD web-portal can be seen as a good practice here (pun intended), because of its ambition to facilitate an open debate. While power will also be a factor in such debates, it at least acknowledges the highly contested, political nature of benchmarks, indicators and rankings.
Notes

1 Think of the gender-related development index and the gender empowerment measure (UN Development Program); the gender equity index (Social Watch); the global gender gap measure (World Economic Forum); the social institutions and gender index (OECD).
2 The EU R&D Scoreboard, the State Aid Scoreboard, the European Innovation Scoreboard and the Internal Market Scoreboard are but a few examples.
3 The Group of Specialists on mainstreaming of the Council of Europe set out to ‘carry out a survey of measures taken and implemented for the mainstreaming of the gender perspective into general policies’ in 1996, just after the launch of this strategy (see Appendix B, Group of Specialists 1998: 80).
4 Women and poverty; Education and training of women; Women and health; Violence against women; Women and armed conflict; Women and the economy; Women in power and decision-making; Institutional mechanisms for the advancement of women; Human rights of women; Women and the media; Women and the environment; The girl-child.
5 The GID index ranks Austria third on violence against women legislation.

Bibliography


