

Performance management of courts and judges: organizational and professional learning versus political accountabilities

Philip M. Langbroek: Director of Montaigne Centre Rule of Law and Judicial Administration and Professor of Justice Administration and Judicial organisation at Utrecht School of Law, Utrecht University, the Netherlands

Rachel I. Dijkstra: Junior Researcher and Lecturer at the Institute for Jurisprudence, Constitutional and Administrative of Utrecht School of Law, the Netherlands

With the cooperation of

Kyana Bozorg Zadeh: Masterstudent and Student Assistant at the Institute for Jurisprudence, Constitutional and Administrative of Utrecht School of Law, the Netherlands

Zübeyir Türk: Masterstudent at Leiden School of law and student assistant at the Institute for Jurisprudence, Constitutional and Administrative of Utrecht School of Law, the Netherlands¹

1. Introduction

Over twenty years of experiences with performance management and measurement in judiciaries have not delivered a one size fits all methodology for assessing and improving the functioning of court organizations and judges. Neither in the world, nor in the EU. Judges work in court organizations. Court organizations, are a part of the state infrastructure², are financed with public money, and, as any public organization, need management and have to account for how the budget is spent.

Judges must be independent from the executive, meaning that they must not be pressured in any way towards a specific outcome in a case, while applying the law in deciding the cases brought before them. Rules of procedure to be applied by the judges and the court, and organization rules for the courts are the product of the legislative and consist of formal and delegated legislation, the latter usually produced by a ministry of justice.

All the same, they are part of the state infrastructure and work in an organization that is dependent on parliament, the legislative and an executive office in terms of organizational rules, budget and rules of procedure. The paradox here is evident; judges must be independent, but effectively they are not in terms of the organizations they work in and in terms of their salaries and the budgeting of the courts. In this context, judges cooperating intensively with court clerks, working in an organization that for its organizational rules (and rules of procedure) effectively depends on cooperation with a ministry of justice and parliamentary votes, are also subject to performance measurement and performance management. Indeed, court operations are monitored and accounted for.

In the debates about court management and court administration, the recurring question is how the interaction between organizational functioning and judges may evolve, both from a judicial impartiality and independence value perspective and a (democratic) accountability perspective. Effectiveness and efficiency are implied here.³ Also in those respects, performance measurement and management for courts and judges is complicated by the

¹ We are grateful for the remarks on an earlier version by the members of the Handle with Care team (Caroline Foulquier, Mátyás Bencze, Agnes Kovács, Francesco Contini, and Petra Pekkanen)!

² Shapiro, Martin, *Courts, a comparative and political analysis*, University of Chicago Press. 1981, p. 3.

³ Marco Fabri and Philip M. Langbroek, (Eds.), *The challenge of change for judicial systems in Europe. Developing a public administration perspective* Amsterdam/Oxford/Washington 2000: IOS Press.

paradox described above. Therefore, a principled question is: is it possible to conduct performance management in a judiciary without putting too much pressure on judges so that content and outcomes of judicial decision making are being influenced? And if so, how?

In this paper, we consider how these questions may be answered. We will first give an inventory and analysis of evaluation methodologies for the functioning of courts, judges and court administrations that are identified in literature. In the different contexts of judicial organizations the interactions between evaluation outcomes and efforts to improve the functioning of courts and judges often are marked by tensions between professional judges, court management, national court administration and national politics. The key to success and failure of policies, organization development and professionalization in judiciaries therefore is most likely to be found in the ways these tensions are mediated and balanced.⁴

Below, we will first describe more extensively the position of courts within the state, in relation to national court administration, and to judicial values and accountability of courts and judges. Second, we describe quality management, and third we will describe performance management methods, based on literature and on reports from Finland, France, Hungary, Italy and the Netherlands, and we discuss those methods in relation to their aims and in relation to judicial values. Fourth, we will give a comparative overview of actual performance management based on the five national reports. We will conclude this report with a discussion of our findings and give some recommendations for the development of performance management in court administration and for performance measurement of courts within the EU.

2. The position of courts and judges in the state and in national court administration

We should make a distinction between courts as organizations and courts as a collection of individual judges. The judiciary is the third state power, and judges are office holders within that power. From a constitutional perspective, judges must be independent from other state-powers, and ought to be impartial in their relations to the parties in the cases they decide. In the continental tradition of the *rechtsstaat*, judicial decision-making is bound by legislation while judicial discretion is (supposed to be) limited.⁵ The judges function in a hierarchy of first instance courts, intermediate courts and highest courts, often with a dichotomy between ordinary (civil and criminal) jurisdictions and administrative jurisdictions. The hierarchy's main purpose is to guarantee to parties and society that the law is administered consistently throughout a country. The judges of the highest courts have the lead in achieving this, on a case by case basis. National legislation arranges for the rules of procedure to be used by the courts and the judges when preparing and deciding and executing cases. Usually there is a division in civil (trade and family), criminal and administrative law. Within these domains there may be further specializations. This is usually referred to as the civil service model.⁶

The national court administration and the courts are responsible for the organization of judicial work. Within the membership of the Council of Europe, most countries have a council for the judiciary with either a role in selection of judges and disciplinary supervision of the judicial profession, or a role in administration of the courts. All countries have a

⁴ Langbroek, Philip and Mirjam Westenberg, *Quality work in four judiciaries*, Justzimanagement Series, Stämpfli Verlag Bern, chapter 7, expected March 2018),

⁵ J.H. MERRYMAN, *the French Deviation*, 1996, *American Journal of Comparative Law* 44, p. 109-119.

⁶ MARY L. VOLCANSEK, *Appointing Judges the European Way*, 34 *Fordham Urban Law Journal*, 2007, p. 368.

ministry of justice that is responsible for legislation in the field of court administration and the judiciary; in some countries, the administration of the courts is a ministerial responsibility (France, Finland). In many countries, the ministry of justice also plays a role in submitting budget proposals for the judiciary and the courts to parliament (Italy, France, Hungary, Netherlands, Finland). They depend on the ministry of justice and the government for primary and secondary legislation, because rules for court administration and rules of procedure need adaptation from time to time.

Hence, effective court administration always requires a close cooperation between the responsible court administration body and the department within a ministry of justice that is responsible for legislation. And usually, budgeting comes with accounts. In other words, the courts and the court administration somehow need to show to the government, parliament and the public what budget they need, how the money will be spent, and how the money has been spent, following general rules for budgeting and accounting.

National legislation arranges the rules of procedure, the judicial map, the jurisdictional organization of courts over topics and territory from first instance to the highest level, and for the budgeting and the financial and organizational accountability of the court organizations towards the political domain, either via a council for the judiciary, a Ministry of Justice (Finland) or a combination of both (Netherlands, Italy, France). In Hungary the Head of the National Judiciary plays a key role in this process. Typically, national court administration and judges participating there, may advise a ministry of justice or a parliament on legislation and secondary and tertiary legislation affecting the functioning of the courts, for example rules of procedure, changes in the judicial map, finances, but also administration rules and changes in legislation that may affect the amount of cases to be filed at courts.

In Europe, except for several Swiss cantons, where judges are elected by popular vote (for example Geneva, Aargau, St. Gallen) judges are appointed by a government or a parliament, following a selection procedure. In this selection procedure, judges and the government cooperate, one way or another, as judges participate in selection committee's, and the government and in some cases the parliament appoints. Usually the government follows the preferences of the judicial selection commissions. The same holds for career decisions regarding judges.

Judicial independence and impartiality demand that the professional discipline of the judiciary is administered by a body that is independent from executive offices. This may be organized in a judicial council consisting of judges, lawyers and prosecutors, like in Italy, France and Portugal, but it may also be kept within the judicial hierarchy, where for example a court president may give a reprimand if a judge misbehaves (Netherlands, Finland). In exceptional circumstances a judge may be removed from office by the highest court or a high judicial council. Anyway an independent court should be able to have the final say in dismissal of a judge.

3. Quality management

3.1. Concept of quality management

The concept of quality management has been thoroughly developed by several well-known international institutions: The International Organization for Standardization, ISO⁷, with its ISO 9000 series for quality management; the European Foundation for Quality Management,

⁷ www.iso.org/iso/home/standards/management-standards/iso_9000.htm.

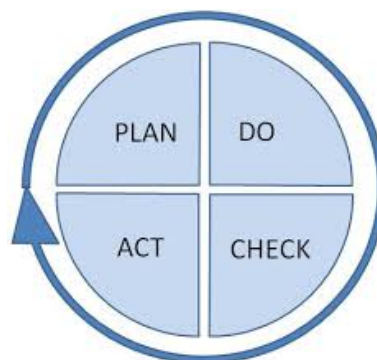
EFQM⁸; and the European Institute of Public Administration, EIPA with the Common Assessment Framework⁹. Each of those organizations has developed a framework for quality management in organizations.

The origin of quality management lies in the product-control approach that started in the Japanese trade and industry in the 1950s: Quality Control. It was a way to ensure that matching products always fit, for example ammunition for guns, tires for cars, and plugs for plug connections. Such products need to have product consistency. Quality control involves standardization of the production process in order to achieve product consistency. This is applicable to many products: e.g. earthenware, such as bowls, pots, cups and plates that need recipes for the baking and glazing processes. To ensure the quality of a product, not only its properties are to be specified beforehand but also the best way to realize these properties, like following a fixed standard recipe.

Continuing along the line of the example of earthenware, hand-made production would follow experience and routine, allowing for some variation in size, weight and color, but within a certain standard range. When scaling up production and developing machines to produce the pottery, the weight, sizes, colors, including their allowed statistical error margins, need to be defined beforehand. Next, checks are necessary to see if these standards are met during as well as after manufacturing. If product standards are not met, the production process has to be adapted; if product standards are met or exceeded, new targets may be set.

This is actually an application of a basic and well-known model for quality management: the quality cycle also known as the Plan-Do-Check-Act cycle (PDCA cycle), or the Demming cycle, named after its commonly presumed creator William Edward Demming who also played an important role in developing quality control in the trade and industry in Japan.¹⁰ The cycle is visually presented below.

Figure 1: Demming's Plan-Do-Check-Act Circle



⁸ www.efqm.org.

⁹ <http://www.eu/caf/>.

¹⁰ The Plan-Do-Check-Act (PDCA) cycle is usually credited to William Edward Demming, who himself however called it the 'Shewhart cycle', referring to the work of Walter Shewhart who developed methods for statistical analysis and quality control. Demming later developed the cycle into the Plan-Do-*Study*-Act (PDSA) cycle, see: DEMMING (1993).

This cycle is also at the basis of the ISO, EFQM and CAF quality management models and the concept of Total Quality Management.

3.2. Total Quality Management (TQM)

After World War II a new management philosophy arose in the Western world, starting in the USA, which was based on the principle of quality control of products and production processes, as described in the previous paragraph, but involving a more comprehensive concept of quality, therefore called 'total' quality management (TQM). It is a broad and systemic approach to managing organizational quality, which became increasingly popular with businesses in the private sector in the 1980s. The keywords associated with TQM are efficiency, effectiveness, economy.

TQM recognizes that customers go to the competition if their product expectations are not met. This resulted in a customer-oriented evaluation of product quality and in a continuous effort to not only meet but to also exceed customers' expectations, through continuous improvement. The leadership of the organisation was acknowledged to have an important role in seeing to continuity, i.e. the cyclical nature of improvement. This continuous improvement not only concerned products, but also the production processes, in which the employees (work teams) were attributed an important role.

The inclusion of organizations' employees in the TQM philosophy reflects the recognition of the concepts of Organization Development, an academic discipline which originated in the behavioural sciences in the 1940s¹¹ and by the 1980s had evolved into a multidisciplinary approach to achieving organizational improvement. This discipline sees organizations as networks of people, and studies how organizations can improve their efficiency and problem-solving ability through their employees.

Essential for organization development is the organization's ability to learn. Argyris and Schön coined the terms 'single-loop' and 'double-loop' learning for organizations.¹² Single-loop learning is about problem-solving by detecting and correcting errors related to fixed goals, values, plans or rules. Double-loop learning goes beyond this: it entails looking inward, for managers as well as employees, to reflect critically and identify the ways their reasoning and behaviour unintentionally contribute to the organisation's problems. In particular, the very way they define and solve problems may actually cause problems. In other words, in double-loop learning, given or chosen goals, values, plans and rules are questioned rather than only operationalized and measured as they are in single-loop learning. Double-loop learning is much more dynamic than single-loop learning.

The classic example to capture this crucial distinction is the thermostat analogy: "When the error detected and corrected permits the organization to carry on its present policies or achieve its present objectives, then this error-and-correction process is single-loop learning. Single-loop learning is like a thermostat that detects when it is too hot or too cold and turns the heating on or off. The thermostat can perform this task because it can receive information (the temperature of the room) and take corrective action. Double-loop learning occurs when error is detected and corrected in ways that involve the modification of an organization's underlying norms, policies and objectives."¹³ A thermostat that asks "Why am I set at 68

¹¹ Its origin was the work of Kurt Lewin on group dynamics in the 1940s.

¹² ARGYRIS/SCHÖN (1974).

¹³ ARGYRIS/SCHÖN (1978), p. 2.

degrees?’’ and then explores whether or not some other temperature might be more economic to achieve the goal of heating the room, would be engaged in double-loop learning.¹⁴

In an organizational context, this means that the monitoring of processes and results is accompanied by reflection on the aim of the organization and interaction with the environment, and, possibly, as a result, by adaptation of aims and standards, in order to achieve a better fit between organization and its environment. This shows that Total Quality Management may reach its purposes if organization aims and results are subject to ongoing evaluation processes. The TQM model was a main driver for the recognition of the concept of Organisation Development.

3.3. Total Quality Management in Courts.

Total Quality Management in Courts has been discussed in the USA, in the Netherlands and Finland, on various levels and it has been applied with various degrees of success. Already in 1994 David Aikman produced a small book on TQM in courts, discussing all the factors that play a role in the success or failure of quality management in this specific type of environment.¹⁵ In this small book, he typically focusses on internal processes in order to improve external relations and the public image of the courts. Typically, he also pays attention to the workforce.

Another USA example of TQM is the international court excellence model, that aims at empowering court management to measure performance in key area’s of the courts and use the measurement outcomes as information for organizing change for improvement. It promotes to use the following key area’s for court improvement:

- Court Leadership and Management
- Court Planning and Policies
- Court Resources (Human, Material and Financial)
- Court Proceedings and Processes
- Client Needs and Satisfaction
- Affordable and Accessible Court Services
- Public Trust and Confidence

The idea is, that a local court organization is able to measure and steer the organization development regarding these subjects.¹⁶ Beneath those themes are values: Equality (Before the Law), Fairness, Impartiality, Independence of Decision Making, Competence, Integrity, Transparency, Accessibility, Timeliness, and Certainty.¹⁷

4. New Public Management (NPM)

New Public Management (NPM) is the management approach that was introduced to public administration in the 1980s. As it also involved quality management, it represents a management ideal for public administration. Until then, for about a century, public administration was based on a type of bureaucracy which had a central role in policy making

¹⁴ ARGYRIS (1991).

¹⁵ AIKMAN, ALEXANDER B. (1994) Total Quality Management in the Courts, a handbook for Judicial Policy Makers and Administrators, National Center for State Courts, Williamsburg.

¹⁶ International Framework for Court excellence 2013, p. 5-13.

¹⁷ Ibidem p. 12

and implementation.¹⁸ Governments and semi-governmental organizations used NPM to modernize the public sector with the intention to make it much more citizen/customer oriented. It consisted of applying management knowledge and experiences from the private sector, including total quality management (TQM), to the public sector. Originating in the Anglo-American area, NPM spread to the Australasia and Europe during the 1980s and 1990s.¹⁹

In the practice of public administration, the introduction of NPM for most organisations represented the first steps in quality management and organisation development. NPM is output-oriented and implies performance measurement, cost management, use of competition and contracts for suppliers, resource allocation and service delivery. In NPM, managers have control responsibility. This means they have to take action if performance measurement results are not in line with set targets. In that sense, they are on top of performance evaluation processes. However, there are differences between private-sector and public-sector organizations. Most notably, private-sector organizations are accountable to themselves or their shareholders mainly, whereas public-sector organizations are accountable to the authorities that may be held directly and politically responsible. The NPM approach to public administration has advocates and it has critics. It has been criticized most for its intra-organizational focus in an increasingly multi-faceted world and for its adherence to the application of outdated private-sector techniques.²⁰ This has resulted in ignoring the linkages between different government services, thus adversely affecting the quality delivered to citizens, and leading to disasters.²¹

It is important to note that ‘checking for performance targets’ may be qualified as quality management, but that for an organization to adapt itself to changing circumstances, it needs more than that. It needs the flexibility and agility to redefine its ‘underlying norms, policies and objectives’ and to change its internal functioning and the interaction with its environment, and therefore to change its standards of performance accordingly. In other words, ‘double loop’ learning within an organization is essential for effective quality management.

For public-sector organizations this entails specific challenges. Quality management is an approach for any organisation to develop itself in interaction with its societal environment, as a result of ongoing evaluation processes. There are, however, marked differences between private-sector and public-sector organizations. Organizations in public administration offer services to users who cannot individually go elsewhere if quality is insufficient: the organizations’ accountabilities are public and political, governed by legal duties. Politics can be very dynamic, resulting in rapid-changing public policy. Consequently, public-sector organizations must be flexible and adaptable. This presupposes an organizational willingness to learn and to be flexible, in terms of action and development, in accordance with public and political accountabilities.

¹⁸ OSBORNE (2006).

¹⁹ A book that inspired many in this context is: D. OSBORNE AND T. GAEBLER, *Reinventing Government, how the entrepreneurial spirit is transforming the public sector*, Boston MA, Addison Wesley. 1993,

²⁰ See for an overview: OSBORNE (2006).

²¹ DUNLEAVY, PATRICK, HELEN MARGETTS, SIMON BASTOW AND JANE TINKLER (2005), "New Public Management Is Dead—Long Live Digital-Era Governance", *Journal of Public Administration Research and Theory* vol. 16, pp. 468, 474-476.

5. Performance management

5.1 Performance management and evaluation

Where quality management is focused on organization development, based on several feedback loops between the (changing) societal environment to the organization, performance management focuses on improving processes and output in terms of efficiency and quality. Where Total Quality Management has evolved into a strategic concept of organization development in interaction with the organizations' stakeholders and with society, performance management aims at enhancing production processes with a view to efficiency and quality. Of course, New Public Management also used performance measurement, especially with a focus on improving services to the public. Performance management is part of a strategic concept for an organization, both with a view to organization development and with a view to accountability. It is important to notice that performance management is based on information provided by performance measurement, a special type of *evaluation*. There may be, but need not be a special focus on relations between the organizations and citizens/customers. It is most likely that performance management in the public sector comes with a purpose: either organization development, human resources development or accountability for performance to the public and to political domains. Usually those purposes are combined.

To date, performance management not only encompasses performance from employees or departments from time to time; it also supposes an ongoing interaction between management and employees about their work and their needs to perform according to standards and policies.

Performance measurement generates the information necessary to conduct performance management. It involves the implementation of measurement devices, for measuring production and for measuring quality. Measuring production requires adequate registries and accurate registration procedures, both for quantitative and qualitative aspects. Even although performance management has always existed in some way or another, to date the buzz word is 'evidence based policies'.²² The idea is that politicians and managers take well informed decisions based on accurate data.

Performance management based on performance measurement is not uncontested, because it may lead to perverse effects in the behavior of employees and to excessive administrative burdens.²³ It takes a lot of time and effort, to have performance based management accepted in an organization.²⁴ In professional organizations, economic focus on production and efficiency may lead professionals away from their basic values.²⁵ Others warn for too much reductionism of reality in performance measurement²⁶, or find that the context where performances are measured make a difference. According to Abma and Noordegraaf, in professional working environments, performance measurement can be harmful, because there

²² VAN DOOREN, WOUTER/BOUCKAERT, GEERT/HALLIGAN, JOHN (2010). Performance management in the public sector. Abingdon: Routledge, p. 43-47.

²³ SEDDON, JOHN Systems Thinking in the Public Sector, Triarchy Press, Axminster 2008.

²⁴ JOHN MAYNE. Challenges and Lessons in Implementing Results-Based Management, Evaluation 2007, issue 13; pp.87-109.

²⁵ VOSSelman, ED (2013) Accounting, Accountability, and Ethics in Public Sector Organizations, toward a Duality Between Instrumental Accountability and Relational Response-Ability, Administration & Society, vol. 48, 5: pp. 602-627.

²⁶ VRIES, MICHIEL S. DE (2010) Performance measurement and the search for best practices, International Review of Administrative Sciences, vol. 76, 2: pp. 313-330.

is too much ambiguity in performances of professionals.²⁷ For judiciaries, this means that there is not a common understanding of what a well-functioning judge is. One of the difficulties also being that judges working in different legal fields may do very different types of work, e.g. trade, versus insolvencies versus family cases. The differences are also in the ways they distinguish between complex and routine cases.

Contini and Mohr have written about performance measurement for courts and judges. They distinguish in the context of courts, four types of judicial accountability, each with its own 'forum'²⁸: managerial (forum: managers, ministries, judicial councils), legal (forum: lawyers, other courts, superior courts), public (forum: media, organised groups, lobbies, individuals), and cooperative accountabilities (forum: quality conference, peer groups). Interestingly, as a forum for managerial accountability they mention not only office managers and judicial councils, but also ministries. This implies that managerial accountability may become political accountability. From a performance management perspective the type of forum for accountability is crucial. Quality management in court organizations generates large amounts of information concerning the performance of the courts and judges, especially where ICT's are used for case management purposes.²⁹ Delivering performance information to local court management is different from delivering performance information to a national court administration authority and to the political domain. Indeed, the latter's scope is much wider than managerial only, it is also political.

Hanne Foss Hansen made an inventory of evaluation methods and their purpose in different circumstances.³⁰ She discerns models according to a focus on:

- Results (goals and effects)
- Systems (performance as a whole)
- Explanatory process model (activity level, implementation problems)
- Economic (cost-effectiveness, cost- efficiency, cost benefit)
- Actor (clients, stakeholders, peers)
- Programme/theory

She combines those evaluation focusses with the aims or purposes of the evaluation at hand: control or learning.³¹ If the goal of the evaluation is to control, it is best to choose a goal-attainment model. Is it about learning, the option would be to choose for a stakeholder model, according to Foss Hanssen. Of course, this is also related to the type of forums. This also has implications for the preferred methodology. Top down control induced evaluations follow a quantitative mode; learning induced evaluations follow a qualitative process oriented mode. Typically, evaluations with a learning aim, have a focus on process; initiatives are often taken by stakeholders and the evaluation is conducted as self-evaluation by peers or by consultants, and the forum consists of peers and stakeholders. Evaluations with a focus on control are organised top-down, and use quantitative measurement methods, when the forum consists of supervisors and policymakers from the political domain. For evaluation of legislation as a

²⁷ T. ABMA. M. NOORDEGRAAF, (2003) Public managers amidst ambiguity. *Evaluation* Vol 9(3): 285–306.

²⁸ CONTINI, F. AND MOHR, R. (2008). *Judicial Evaluation, Traditions, Innovations and Proposals for Measuring the Quality of Court Performance*. Saarbrücken: VDM-Verlag.

²⁹ GAR YEIN NG, MARCO VELICOGNA AND CRISTINA DALLARA, *Monitoring and Evaluation of Court System: A Comparative Study*, CEPEJ Strasbourg, 2008, p. 47-48, point out that a lot of data are collected in Council of Europe Memberstates, but that they are based on different standards, and therefore not fit for comparisons.

³⁰ HANNE FOSS HANSEN, *Choosing Evaluation Models, A Discussion on Evaluation Design, Evaluation 2005*. Vol 11(4): 447–462.

³¹ *Ibidem*, p. 452.

product of democratic decision-making, Hansen asserts that the top-down result oriented control model is appropriate, whereas when professional groups are involved whose autonomy should be respected, a peer review model should be appropriate.³²

Foss Hanssen speaks in her article about three rationalities when making a choice for an evaluation model.³³ First of all, one can choose for a certain evaluation model because of the purpose of the evaluation to be performed. Secondly, an evaluation model can be chosen based on the characteristics of the evaluated program or organization, and thirdly, it can be based on the characteristics of the problem that the policy program or organization wishes to solve. There may also be multiple rationalities in an evaluation. In accordance with Foss Hansen's model, that multiplicity should be reflected in the methods applied for those different purposes.

Based on the literature described, performance management and its related evaluation processes may not be applied unconditionally for courts and judges, because courts as organizations and judges as professionals combine routine work and non-routine work with multiple sided interactions, and because courts interact with parties and lawyers of a highly differentiated character.³⁴ They need enough time and leeway to decide how to deal with a case: routinely or expertly, both from skills and required knowledge perspectives. Nevertheless, also professional judges can be asked to adapt their work and organization to changing circumstances, also when technological and societal developments change.³⁵ The evaluation methodology of Foss Hanssen may be able to solve the tensions between judicial independence and accountability, and preserve judicial professional autonomy while also allowing for public and political accountability, because it departs from different evaluation purposes. It allows for flexible approaches, depending on evaluation purposes, recognizing both the special position of professionals in organizations, and the need to inform the political domain about results and the ways in which public money has been spent, in representative democracies.

5.2 Performance Measurement and Evaluation Tools for Courts and Judges

During the past decades, quite some literature has been published on how to understand the position of courts and judges in the state organization, from a non-legal perspective. This literature has focused on reconciling constitutional demands for judicial independence and impartiality with accountability of the courts and the judiciary from a societal and democratic process point of view, and may have contributed to the development of tools for evaluations of courts and judges.³⁶ We will not discuss this literature here; instead we will describe

³² Ibidem, p. 451-452.

³³ Ibidem, p. 451

³⁴ ABMA and NOORDERGRAAF 2003, p. 291.

³⁵ MIRKO NOORDEGRAAF, Reconfiguring Professional Work, Changing Forms of Professionalism in Public Services, Administration & Society, vol. 48, 7: pp. 783-810, 2013.

³⁶ KLAUS RÖHL, Gerichtsverwaltung und Court Management in den USA, Köln, Bundesanzeiger Verlagsgesellschaft 1993; MARCO FABRI AND PHILIP LANGBROEK (eds). The Challenge of Change for Judicial Systems, developing a Public Administration Perspective, a collection (IOS Press Amsterdam/ Washington DC. EGPA yearbook 2000; MARIE-LUC CAVROIS, HUBERT DALLE, JEAN-PAUL JEAN, la Qualité de Justice, La Documentation Française, Paris 2002 ; MARCO FABRI, JEAN- PAUL JEAN, PHILIP LANGBROEK. HELENE PAULIAT (eds. and research directors. L'administration de la justice en Europe et l'évaluation de sa qualité. Collection Grands colloques", Montchrestien, Editions juridiques associées, Paris 2005, (eds. and research directors.); FABIAN WITTECK, Die Verwaltung der Dritten Gewalt, Mohr Siebeck, Tübingen, 2005 ; GAR YEIN NG, Quality of Judicial Organisation and Checks and Balances, March 2007; FRANCESCO CONTINI AND RICK MOHR, Judicial Evaluation, Traditions, Innovations and Proposals for Measuring the Quality of Court Performance. Saarbrücken: VDM-Verlag 2008; BENOIT FRYDMAN ET EMMANUEL JEULAND (directions), Le nouveau Management

currently developed tools for evaluation of courts and judges, and refer to literature when relevant for our descriptions.

Measurement tools for courts and judiciaries have been developed in various kinds and sizes throughout the past decades. We refer to the work of the European Commission for the Efficiency of Justice (CEPEJ) of the Council of Europe³⁷, the International Framework for Court Excellence³⁸, to the work of the Finnish Rovaniemi Court of Appeal³⁹ and the EU-Justice Scoreboard⁴⁰, to the Lower Saxony Benchmark project, and to the Internal and External Dialogue of Marie Hagsgård.⁴¹ We reiterate here, that it matters what and whose performances are measured and how the performances are reported and to whom they are reported. Ownership of management information on performances of courts and judges is an issue, from a judicial independence and professional autonomy perspective.⁴² Purposes of gathering and sharing information can be very different: professional learning, organisation development & local management, or informing the public and policymakers. For each of these purposes providing different types of information are the result of balancing judicial independence and professional autonomy on the one hand and the purpose of the information provision on the other.

First of all, the CEPEJ has engaged since 2006 into publishing biannual reports showing the state of affairs of the judicial systems in the member states of the Council of Europe: *European judicial systems Efficiency and quality of justice*.⁴³ Those reports compare for example the number of categories of cases, average times of specific court proceedings, the number of judges per 100.000 inhabitants, the budget spent on the courts, judicial salaries, judicial independence, the position of prosecutors and so on.⁴⁴ The repetition shows the changes per country, and also makes comparison between the countries possible over time. The tools still have flaws, because the question list is answered by civil servants on a national level and the reliability of the data presented is not always clear.⁴⁵ The quality of justice systems evaluation reports typically is an information tool for national policymakers in the justice field.

The CEPEJ does much more, however, also with a focus to the management of individual courts. For example, the CEPEJ has been working on *evaluation methods for courts*.

de la justice et l'Indépendance des Juges, Paris, Dallos, 2011; FRANCESCO CONTINI, AND SHARYN ROACH ANLEU, AND DAVID ROTTMAN (editors), *Evaluating Judicial Performance*. Oñati Socio-Legal Series, Vol. 4, No. 5, 2014; ANDREAS LIENHARD AND DANIEL KETTIGER, *The Judiciary between Management and the Rule of Law*, Bern, Stämpfli 2016.

³⁷ www.coe.int/cepej: European judicial systems Efficiency and quality of justice, CEPEJ STUDIES No. 23.

³⁸ <http://www.courtexcellence.com/>.

³⁹ *How to Assess Quality in the Courts?* Rovaniemi, 2006

⁴⁰ http://ec.europa.eu/justice/effective-justice/scoreboard/index_en.htm (last accessed 6 December 2017).

⁴¹ HAGSGÅRD, M., (2008). Internal and External Dialogue: A Method for Quality Court Management. *International Journal for Court Administration*. 1(2), pp.10–18. DOI: <http://doi.org/10.18352/ijca.98>; HAGSGÅRD, MARIE B, Internal and External Dialogue: A Swedish Approach to Quality Work in Courts (December 21, 2014). Oñati Socio-Legal Series, Vol. 4, No. 5, 2014. Available at SSRN: <https://ssrn.com/abstract=2541449>.

⁴² LANGBROEK, PHILIP. (Ed.) (2010). Quality management in courts and in the judicial organisations in 8 Council of Europe Member States: A qualitative inventory to hypothesise factors for success or failure. Council of Europe, European Commission for the Efficiency of Justice: CEPEJ (2010). p. 13.

⁴³ http://www.coe.int/t/dghl/cooperation/cepej/evaluation/default_en.asp.

⁴⁴ The youngest scheme for evaluating judicial systems 2016-2018 cycle counts 62 pages and 208 subjects.

⁴⁵ JOHNSEN, J.T. (2012). "The European Commission for the efficiency of justice (CEPEJ). Reforming European justice systems – 'Mission impossible?'" *International Journal for Court Administration*, Special issue, 1–19. Available at: http://www.iaca.ws/files/ijca_2012_special_issue/ijca-2012-special_issue.pdf.

Examples are the court user survey by Jean Paul Jean and Helène Jorry and Martial Pasquier.⁴⁶ Furthermore the CEPEJ recently published a document prepared by Fabio Bartolomeo from the Italian Ministry of Justice.⁴⁷ This document proposes a set of quality indicators at the court level in order to facilitate court management both at the local and at the national level. It explicitly distinguishes measurement for inspection and performance measurement for management.⁴⁸ Last, but not least important, the CEPEJ has set out since 2002 on improving timeliness of court proceedings in the SATURN project.⁴⁹ Here the CEPEJ has developed guidelines for judges⁵⁰ and for court management.⁵¹ The focus is almost exclusively on the court level.

The *International Framework for Court Excellence* also takes the autonomy of court organisations for granted. It starts from a set of values and then instructs court managers on how to measure performance on those values. These values typically are:

- Equality before the law
- Fairness
- Impartiality
- Independence of decision-making
- Competence
- Integrity
- Transparency
- Accessibility
- Timeliness
- Certainty⁵²

The International Consortium for Court Excellence gives also advice on how to operationalize these values in courts, so that performance in realizing those values can be measured, with a view to organization development and improving performance to a desired level. In an explanation on performance measurement and management it is stressed that these values not only apply to single courts but also to justice systems.⁵³ Implicitly the scope of the IFCE is thus extended from a mix of professional learning, management information and organization development, towards informing policymakers with a view to policy development in the justice field.

The *Rovaniemi Court of Appeal Quality Benchmark system*⁵⁴ is a tool developed to improve the services of the court to its customers, based on the CAF quality management system. Measuring performance on a set of pre-set values is essential with a view to see if interventions were effective. The points of attention are:

⁴⁶CEPEJ (2016) 15 HANDBOOK FOR CONDUCTING SATISFACTION SURVEYS AIMED AT COURT USERS IN COUNCIL OF EUROPE MEMBER STATES.

⁴⁷CEPEJ (2016)12, Measuring the quality of justice, Document prepared by the CEPEJ-GT-QUAL on the basis of the preparatory work of Mr Fabio BARTOLOMEO (CEPEJ member, Italy), 7 December 2016.

⁴⁸Ibidem p. 5.

⁴⁹http://www.coe.int/t/dghl/cooperation/cepej/Delais/default_en.asp.

⁵⁰CEPEJ (2014) 16, REVISED SATURN GUIDELINES FOR JUDICIAL TIME MANAGEMENT.

⁵¹CEPEJ (2015) 18, IMPLEMENTING THE SATURN TIME MANAGEMENT TOOLS IN COURTS, A GUIDE.

⁵²International Framework for Court Excellence, 2013, p. 3.

⁵³INGO KEILITZ, DANIEL HALL, LAURENCE GLANFIELD, Global Measures of Court Performance, Second Edition, 2017.

⁵⁴How to Assess Quality in the Courts? Rovaniemi, 2006.

- The process
- The decision
- Treatment of the parties and the public
- Promptness of the proceedings
- Competence and Professional Skills of the Judge
- Organisation and Management of Adjudication

The focus here is on judicial performance and the aim of the Rovaniemi tool is professional learning.⁵⁵

In 2002, the courts in the German Federal State of Lower Saxony pioneered with *comparison circles for benchmarking*, starting with two pilots, one with six district courts and one with eight district courts (Amtsgerichte im Leistungsvergleich, AgiL).⁵⁶ This benchmarking project consisted of a zero measurement of court performance in a certain field, a workshop where practices were inventoried and compared, and another performance measurement after two years. This then has led to standard protocols for all the courts at the same level.⁵⁷

The enthusiasm of the participants and the success of the practical results combined with the support from the Ministry of Justice convinced all other district courts of Lower Saxony to take part in the project. This was later followed by comparison circles for the regional courts (Landgerichte im Vergleich, LiVe). Because larger courts have other issues to deal with than the smaller courts, separate comparison circles were started for the large district courts and also for the large regional courts.⁵⁸

The benchmarking projects were initiated by and conducted under the supervision of the three higher regional courts (Oberlandesgerichte). Other German federal states adopted the benchmarking method of Lower Saxony as well, and since 2007, the higher regional courts themselves also take part in nationwide benchmarking (Oberlandesgerichte im Vergleich, OliVe).⁵⁹ Clearly, those comparison circles have a focus on professional learning from best practices.

In Sweden and Norway, the Internal and external Dialogue has been used to organise feedback for court staff and judges, first, with a view to their cooperation and upholding judicial values. Second, to ask for feedback to judges and court staff from court users, advocates and so on. Dialogues are organized by the local court organization, the initiative is bottom-up. As the focus here is entirely on professional learning and organization development, the method is qualitative.⁶⁰ In order to come to a joint analysis of the situation and to an agreement of what should be done to improve the functioning of the courts,

⁵⁵ For more information, see paragraph 3.1 in the report on Handle with Care National analysis - Finland

⁵⁶ MATHIAS VOLKER, Qualitätsmanagement in der ordentlichen Gerichtsbarkeit der niedersächsischen Justiz. Projekte AGiL (Amtsgerichte im Leistungsvergleich), LiVe (Landgerichte im Vergleich): Ausblick zum Qualitätsmanagement der Oberlandesgerichte 5. Auflage März 2007 Mathias Volker, Oberlandesgericht Celle. Download from: <http://www.oberlandesgericht-oldenburg.niedersachsen.de/aktuelles/reformprojekte/qualitaetsmanagement/79958.html>.

⁵⁷ LANGBROEK, PHILIP. (Ed.) (2010). p. 78.

⁵⁸ VOLKER, MATHIAS, MICHAEL KALDE AND RALF-GÜNTHER LÜPKES, Qualitätsmanagement der Oberlandesgerichte. Deutsche Richterzeitung, Oktober 2008, pp. 269-271.

⁵⁹ <http://www.oberlandesgericht-oldenburg.niedersachsen.de/aktuelles/reformprojekte/qualitaetsmanagement/79958.html>; LANGBROEK 2010, p. 119-130.

⁶⁰ HAGSGÅRD 2014.

cooperation between court staff and judges is necessary. The president of the court has to fulfil a leading role here, to persuade both judges and court staff to leave their comfort zone.⁶¹

A similar tool has been developed in the Netherlands, with the so called ‘mirror’-meetings, where court users can discuss amongst themselves their experiences with the court and the judges, while the judges and court staff are listening.⁶²

Not long ago, more than ten years after the creation of the CEPEJ by the Council of Europe, the EU Justice Scoreboard was introduced by the EU Commission.⁶³ It is presented as a statistical information tool on the efficiency of member states’ judicial systems.⁶⁴ And as an information tool its announcement in 2013 carries the explicit messages that courts in EU member states also function as EU courts, are important for mutual trust in the EU, and that courts also have a crucial role in economic stability:

*Citizens, businesses, judges and authorities are expected to trust, respect, recognize or execute decisions taken by the justice system in another Member State*⁶⁵

This makes sense, as mutual trust has become a legal obligation for courts in the EU in transnational cooperation in the fields of criminal law (for example, European arrest and evidence warrants) and transnational execution of civil judgments. The program has been established by the Commission in an open dialogue with many authoritative policy makers from different Member States in order to achieve a safe (secure) investment climate, to stimulate economic growth and to maintain the rule of law.⁶⁶ The focus is on the efficiency, quality and independence of the judiciary by means of indicators:

- The length of the proceedings expresses the time (in days);
- The clearance rate (the ratio of the number of resolved cases over the number of incoming cases in a year/month)
- The number of pending cases expresses the number of cases that still have to be treated at the start of a period (e.g. a year)
- The monitoring and evaluation of courts' activities by means of performance indicators and standards
- The perception of the independence and impartiality of the judiciary by the public.
- ICT systems for courts: the availability of ICT systems for registration and management of cases, and for communication and information exchange between the courts and their environment
- Alternative Dispute Resolution methods (ADR)
- Training of judges: initial training and continuous training

⁶¹ Ibidem p. 1008-1109.

⁶² JITTA MIEDEMA, ABELTJE HOOGENKAMP, CORA VAN STEENDEREN KOORNNEEF, MARIA MUL (2015), Spiegelbijekomsten, de stand van zaken na 5 jaar, Feedback vanuit de relevante buitenwereld, Tijdschrift voor de Rechterlijke Macht, p.315-320.

⁶³ The EU Justice Scoreboard, A tool to promote effective justice and growth, Brussels, 27.3.2013 COM (2013) 160 final.

The 2015 EU justice scoreboard, COM(2015) 116 final.

⁶⁴ ADRIANI DORI, The EU Justice Scoreboard - Judicial Evaluation as a New Governance Tool, MPILux Working Paper 2, 2015, p. 6.

⁶⁵ The EU Justice Scoreboard, A tool to promote effective justice and growth, Brussels, 27.3.2013 COM (2013) 160 final, p. 1-2.

⁶⁶ ADRIANI DORI, 2015, p. 13.

- Resources: the budget for courts, the number of judges and lawyers provide information on the resources used in the justice systems.⁶⁷

Many data are collected and assessed against these indicators.

*“The Scoreboard provides a comparative overview of the quality, independence and efficiency of national justice systems and helps Member States to improve the effectiveness of their national justice systems. This makes it easier to identify shortcomings and best practices and to keep track of challenges and progress. In the context of the European semester, country-specific assessments are carried out through bilateral dialogue with the national authorities and stakeholders concerned. This assessment takes into account the particularities of the legal system and the context of the Member States concerned. It may lead to the Commission proposing to the Council to adopt country-specific recommendations on the improvement of national justice systems.”*⁶⁸

Also the perception of the public of judicial independence is being monitored. In conclusion, the Justice Scoreboard is a measuring device of the European Commission to inform policymakers about the performance of justice organisations and best practices within the EU member states, with the intention that national policymakers in the justice field try to improve their justice systems in fields with shortcomings. In that sense, it is also both a comparative instrument and an instrument of control.

Within the EU-context, a positivist, quantitative approach is dominant in policy evaluations.⁶⁹ EU policy makers do not have a lot of control over policy implementation in the different member states. Hence, hierarchical evaluation methods are most likely to deliver results that may be comparable cross national borders. Monitoring typically belongs to the evaluations with a view to control. Within the EU, monitoring typically serves comparative purposes. For the European Commission, “Quantitative data is suitable and convincing, allowing for easy aggregations, comparisons, and generalizations.”⁷⁰ This explains why qualitative, constructivist approaches for EU policy evaluations are scarce, and it shows that the Justice Scoreboard fits the EU evaluation pattern.

Apart from performance information for policy development (and for informing politicians and the general public) there may be a need for information for different purposes: professional learning, local management information with a view to organization development and performance information on an individual level for initial appointment and promotion decisions, upholding judicial values, and legal accountability concerning a courts’ judgments.

6. Inventory of evaluation subjects, methods and actors in the National Reports

Based on an inventory and qualification of performance evaluations in the national reports, we try to come to a conclusion and discussion of possible ways to measure court and judicial

⁶⁷ The EU Justice Scoreboard, A tool to promote effective justice and growth, Brussels, 27.3.2013 COM (2013) 160 final, p. 4-5.

⁶⁸ Ibidem p. 3.

⁶⁹ JULIAN HOERNER AND PAUL STEPHENSON, Theoretical Perspectives on Approaches to Policy Evaluation in the EU: the Case of Cohesion Policy, *Public Administration Vol. 90, No. 3, 2012 (699–715)*.

⁷⁰ HOERNER AND STEPHENSON 2012, p. 712.

performance in combination with enhancing judicial independence and impartiality. So far, we distinguished between performance measurement for evaluation of court work with a view to:

- Professional learning
- Upholding judicial values
- Human Resources Management
- Quality of decision making, consistency of case law, timeliness
- Court management and organization development
- Accounting towards the general public and for policy development

The national reports show several methods of evaluation that apply to different subjects of evaluation. All national reports distinguish between the evaluation of judges and the evaluation of the courts. The following highlights the evaluation mechanisms in place in subsequently Finland, Hungary, Italy, France, and The Netherlands.

6.1. Measuring performance with a view to Professional learning.

Judicial main tasks are to hear cases and to deliver judgments. Conducting a court hearing and delivering judgments implies very different skills: communicating with parties and their councils, managing the case and the exchanges of documents and planning a hearing from a time perspective, writing a judgment alone or in interaction with judicial colleagues and court clerks. This presupposes a high level of expertise on the rules of procedure and the law that is applicable in a case, including on the development of case law and on other legal and societal developments, including expertise on the subject matter. The perspective of the parties on the case is highly relevant, not only to inform the judge, but also from a procedural justice perspective, indicating that judicial behaviour does matter very much for the acceptance of the judgement by the parties. Comprehensibility of judgment texts is also a topic in explaining the legitimacy of judgments. Monitoring techniques here may be peer review on the judge level and on the team level, and also evaluating interactions with stakeholders and court users is possible.

Appropriate methods are:

- Exchanges between judges, court staff and stakeholders on the court level (Hagsgaard⁷¹) Stakeholders at the court level are the prosecutions office, the police, lawyers, parties, bailiffs, the youth protection agency, mediators and the press;
- Case law review (also concerning reversals in appeal);
- Mirror sessions (Miedema and Mul⁷²);
- Individual peer review (court hearings, judgments);
- Court user satisfaction surveys and discussions on actions to be taken, because of the outcomes;
- Judicial participation in training courses (in order to keep their knowledge and skills up to date).

Such methods are used more in Finland and the Netherlands, compared with Italy, France and Hungary. But also there, court user satisfaction surveys are locally implemented, sometimes with national coordination. Peer review on case law is a point of attention everywhere, but not implemented frequently in Italy and France at the local level. Training courses are organised

⁷¹ HAGSGÅRD, M., (2008, 2014)

⁷² JITTA MIEDEMA, ABELTJE HOOGENKAMP, CORA VAN STEENDEREN KOORNNEEF, MARIA MUL, 2015.

everywhere, and it shows that national court administrations go through a lot of efforts to stimulate judges to follow courses.

Conditions here are, that judges and judicial management are in charge. Within the available time judges can shape their own professional space by participating in those activities. From a performance measurement perspective, registrations can be designed that for example show how many judges of a court participated in what activities and when. From a performance and quality management perspective, standards should be set for the registration, with minimum performance requirements. For example, *each court should do a court user satisfaction survey at least once every three years. Or: at least 50% of judges should spend 30 hours per year on judicial training courses.*

If measurement shows inadequate activities, the courts' management or the national court administration should analyse its causes, and take action. A question is, in how far this management information should be reported to the general public and the political domain.

6.2. Registration of judicial performance with a view to upholding judicial values

Judicial values are a mix of constitutional, professional and economic values. Judges are to be independent and impartial; they need to work with their peers, and with representatives of parties, with the parties themselves, with court clerks. To enhance their role of judges, they must behave impeccable. Furthermore, they need to be well advanced in their field, knowledge wise, and also apply the law in a consistent way. Last but not least important, they should work efficiently and organise the management of their caseload in such a way that delays are an exception, regardless the agendas of the parties. Timeliness of judgments and quality of judgments is an issue everywhere.

Upholding judicial values does touch upon professional judicial ethics.

A question is, what methods are used to uphold judicial values. The answer is: judicial values can be upheld in many ways, also beyond control mechanisms and management information. Usually they are ingrained into judges during their training on the job, as a part of the professional socialisation process. It may help to formulate those norms explicitly, for example in a code of conduct. Upholding judicial values is not only an issue of the individual judge's conscience. For example, timeliness of judgement is a judicial value, but for case management, upholding this value is also a matter for the court or court division. We think the following mechanisms for registration and feed-back are in place to uphold judicial values:

- Complaints procedures; usually the president of the court is in charge of them. Complaint procedures are not about the case content, but usually about what a judge did or said before or during case management and hearing;
- Challenging judges, before or during proceedings, because there is an (appearance of) bias towards one of the parties in a case;
- Judicial Personnel Supervision by a judicial council, or other independent body where the majority consists of professional judges;
- The possibility of disciplinary sanctions, with as ultimate sanction removal from office (with an eventual check by a court of independent judges); Professional Peer review;
- Professional Peer review;
- Appeal to a higher court with a case, if legal argumentation about judicial bias or prejudice was not effective in first instance.

It makes a difference in what context those activities take place. The hierarchy of checks on judicial behaviour as in Italy and France is different from the working relations between judges and the courts' boards in Finland and the Netherlands. When the professional norms

are clearly formulated, it can also make sense to register the activities in this regard, and show the outcomes on the court and at the national level and what action has been taken. From an organizational point of view, it is always interesting to know what a court or a national court administration actually did with the outcomes of quarterly or annual registrations, and what happened with complaints.

6.3. Measuring performance with a view to Human Resources Management (also: employee satisfaction survey)

The national reports pay a lot of attention to the professional evaluation of judges. Those evaluations focus on individual productivity, timeliness and quality of judgments, expertise, conduct in the court room, case management skills, and work attitude. The methods do differ considerably.

Italy employs a system of evaluating judges every four years and looks at professional skills (e.g. legal knowledge), productivity (e.g. number of cases dealt with), diligence (e.g. punctuality during hearings), and commitment (e.g. following extra trainings). In Italy, individual judicial evaluation is quite a bureaucratic process, with statistics on the productivity of the judge, evaluation reports by a judge's superiors and the involvement of a regional Judicial Council. The Council for the Judiciary – a national entity – coordinates this process and takes the final decision on promotion or salary increase of the judge.

In Hungary the assessment is done on a regular basis by the Head of the competent division or a Judge appointed by her. The regular evaluation of judges is designed to check the judicial skills of the judges, including quantitative aspects (e.g. number of trial days), and qualitative aspects (e.g. thoroughness). Judges are assessed in the third and sixth year after their appointment and every 8 years following it. The Judge is evaluated through an assessment, which includes an analysis of a certain number of final judgments rendered by the judge, the annual professional activity of the judge, and the opinion of the Head of the division competent in the legal area where the assessed judge works, among other elements.

In France the immediate supervisors of a Judge also have a considerable role when it comes to measuring the performance of a Judge. They conduct individual evaluations every two years and incorporate a professional interview between the judge and the Head of Court (usually the direct supervisor), an overview of the Judge's activities and trainings and a form filled in by the Head of Court on aspects such as the vocational skills and analytical ability.

In Finland a Judge has yearly development meetings with his or her supervisor and is also subject to a statistical check regarding workload.

The Netherlands do not have such a clear guideline on how frequently a judge has to be evaluated. As a matter of policy, judges have annual evaluation conversations with a managing judge. This is much more a professional conversation than a performance assessment. Even so, severely underperforming judges can be reprimanded or, in extreme cases even be dismissed by the Supreme Court. This actually never happens, as such judges usually are pressured in a friendly way to leave their position voluntarily. The Dutch system does have methods such as peer review, co-reading, and mandatory permanent education in order to safeguard expertise and legal quality. Judicial personnel policies are not very transparent however, the presidents of the courts play a major advisory role in promotion of judges.

Measuring individual judicial performance is a very sensitive issue, especially with a view to promotion decisions and judicial independence, also within the court. On the other hand, a rational approach is also necessary to select the best judges for promotion. Then the fairness and (internal) transparency of such assessment and selection processes is important, as preferences of executive office holders related to the content of past decisions of judges should not play any role. Questions here are in how far quantitative approaches or qualitative approaches to assessment or combinations of both are to be preferred. Anyway for the assessments judges – peers – should be responsible, regardless who takes the decisions to (re)appoint a judge in a higher judicial office (usually the government).

Apart from individual evaluations, there is the method of employee satisfaction surveys. These are part of the quality system in the Netherland, are used in Finland, and sometimes in France, but not in Italy and only sporadically Hungary. These surveys are an instrument that may give some insight for the management in the extent to which employees do engage in the work of the organisation, support its goals and feel recognized by their superiors.

Last but not least, it makes sense to measure the influx of cases and output of cases. This may lead to changes in allocation of judges and or court staff, or take measures that persons from other courts are (temporarily) moved to a court with a large influx of cases. Given the development of e-filing and electronic communications the transfer of cases to a court with lesser influx of cases may be an alternative consequence, also presuming such measures are in no way related to putting judges under pressure for career purposes.

Monitoring the production, productivity and timeliness on the court level, on the national level and by case category is also very useful from a HRM perspective, as changes in caseloads may be connected to staff numbers and numbers of judges. On a national level, this presupposes common definitions and standardized registration methods. Standards should be defined when a case starts in a court and when a case leaves the court. It presupposes also adequate insight in the actual amount of personnel staff and judges in the courts, and in the amounts of money spent. This is easier said than done!

6.4. Measuring performance with a view to Management and organisation development

Every organisation somehow needs to monitor the flows of money and of the services/products delivered. For organisations in the public sector, this is not different, and this also applies to the courts.

It is possible to measure (average) productivity and (average) timeliness at different organizational levels from the court level to the individual judge-level, and different teams or departments in between, in a certain time period. Furthermore, it is possible to monitor court user satisfaction and also employee satisfaction. Organisations function in a context, and therefore the perception of the organisation by people in that context are important for receiving feedback. This may also involve periodic peer review: review of the functioning of a court by colleagues from other courts, and by some experts in for example business administration or public administration. It may also be handy to involve journalists. The tools developed by the CEPEJ apply on the court level and on the national level. However, as we have indicated above, this involves choices on what data should be shared between the local level and the national level.

Performance management and organisation development also require that for the measurements, certain targets will be set. Meeting those targets or not meeting those targets – and their explanation is most useful for management purposes.

The basic information concerns: court user and employee satisfaction surveys; production, production costs, productivity and timeliness of court decisions; number of successful challenges of judges; number of cases with exceptional long delays in a certain timespan.

We do not approve of measuring the number of reversals in appeal. That information belongs to the judicial professional space. That indicator does not say much about the quality of judgments or about the professionalism of a judge, because reversals in appeal are also about differences of opinion on technical legal issues.

Usually, total quality management systems are very labour intensive, and in practice they have not been very successful (Norway, the Netherlands). Benchmarking based on local information, can be used by courts to share and apply best practices (Lower Saxony), or professional learning and organisation development (Rovaniemi) – and those have been successful.

In Italy court evaluation focuses on effectiveness and efficiency and aims at managing the courts accordingly. France focuses on speed of justice and productivity. Finland emphasizes productivity, economic efficiency, and timeliness, but adds also qualitative targets such as customer service. The focus on quality of the courts in Hungary also has a strong quantitative emphasis. The performance indicators for the courts are (1) incoming cases, (2) resolved cases, (3) backlog, (4) timeliness (length of proceedings), (5) workload, and (6) appeal ratio. However, practical difficulties can hinder the realization of targets. The Italian Judiciary does not have a workload measurement system (or a standard cost system) that clarifies the average working time needed to define a case (or its average cost). Due to this lacuna, all the planning is based on caseload measures that are just (rough) proxies of the workload of the unit. The Netherlands do have a very specific workload defining system that relates to each type of case.

It is important to notice that quantitative methods to measure performance on the level of a court or a court division usually has a strong control focus. Court managers need information about performance-to be able to take care of the well-functioning of the court organization. This does not need to be a complete overview of everything that can be measured. Organization development, or adapting work processes to new developments, usually are not a quantifiable process. For management purposes the required quantitative information needs to be limited, also in order to maintain enough professional space for the judges. Even so, candid communication between judges and management about ongoing concerns is necessary, and judges need to have enough organizational space in terms of time to focus on the content of their work. This means that it may also be useful to set standards for organization and management related activities for judges. An entirely different matter is, what information on the outcomes of these measurements, should be shared with the general public and the political domain.

6.5. Measuring performance with a view to accounting for public money spent (reporting to politics and to the general public)

What are key topics for reporting to the public and to politics about the performances of courts and judges? What data should be in the annual reports?

Judicial organisations as part of the state organisation are full of paradoxes here. Judges should be independent, but in representative democracies, parliament is the ultimate authority with democratic legitimacy. Parliament authorizes legislation that judges should apply, the legislation for the judicial organization and the budget. Accountability and control are natural elements that are part of such systems. Reconciliation of accountability and judicial

independence is organized by putting judicial professionals in buffering structures, like an administrative council for the judiciary or a the management board of a court, together with management professionals. Another variant may be a separation of organisational responsibility via a ministry of justice and professional supervision via a council for the judiciary. Usually detached judges then take part in the work of the department at the ministry of justice. Legislation makes them responsible for the organisational functioning of the courts, and for the upholding of judicial values respectively. Such organisational patterns are normal within constitutional legal states. But that leaves the question open how the balance between informing policymakers, politicians and the general public with view to democratic control and preserving judicial independence should be struck. As the experience is, that also judges are sensitive to the normal managerial manipulation in organisations, we take a conservative stand in favour of judicial independence.

The first issue here is what information should be reported and the second issue here is at what level of aggregation the information should be reported.

Here we make a distinction between three domains: the political domain, the organisation and management domain and the professional judicial domain. The political domain should not know more than necessary to issue legislation and decide on the budget. The organisation and management domain should not know more than necessary to help courts and judges function impartially and independently, effectively and efficient from an organisation perspective. Because organisation and professionals interact, organisation and management should be dominated by judicial professionals. This also means that the information gathered by the organisation and management domain in judiciaries need not be shared with the political domain.

So for hierarchical control, reporting in numbers about performance is indicated. Regardless the possibility to develop quantitative tools to measure qualitative aspects of the functioning of a judiciary, we take the stand that this should not be done for reporting to the political domain where the subject of such measurements belongs to the professional judicial domain: knowledge management, consistency of judging, reversal in appeal, writing judgements, cooperation between judges and court clerks, peer review, and so on. What can be delivered are quantitative data on production, productivity and timeliness. The data may be presented per legal domain: trade, family, administrative, criminal and so on.

For this subject, we are aware that in some countries, like Hungary and Poland, a severe political distrust in the judiciary prevails, with efforts to achieve a strict control on appointments and on judicial performance also for the constitutional courts. In this report we do not go into the debate about the reforms in those countries. We just refer to our conservative stand that the political domain need not know everything that goes on in a judicial organisation and a court organisation, and responsibilities for management and organisation should be left with institutions like a council for the judiciary or the board of a court, trusting that they will act responsibly, also in light of emerging new circumstances.

One of the most important issues to be addressed here is the level of aggregation at which should be reported. This is not something to be taken lightly. Considering the level of IT automation in courts in many EU-member states, it is not difficult to measure individual performances of judges and court clerks, in terms of number of cases, number of judgments, timeliness of judgments and so on. We take the position that the level of the numbers reported on should be aggregated at the level of the individual court department (for example trade, criminal, family, administrative), so that court departments can be compared nation-wide, or that only aggregated numbers per legal field nation-wide should be presented. The question is

what a ministry of justice and members of parliament will do with detailed information, also in relation to the judiciary.

Of course, organization development in the judiciary may be reported upon on a general level to a ministry and to parliament. The data provided by the CEPEJ judicial systems report are on a national aggregate level. And so are the Justice Scoreboard data. Several countries publish information on the court level, so that comparisons between courts become possible. This may cause questions for explanations of differences in performances between courts. Because courts are very complex organizations, such explanations may be very difficult to accomplish accurately based on quantitative data only. We therefore take the stand that such comparisons do not belong to the political domain, if necessary at all, they belong to the responsibility of the management and organisation domain of judiciaries.

The main method to report to politicians and the general public is reflected in the yearly target negotiations that are being held between courts and the Judicial Council and/or Ministry of Justice and/or Parliament. Finland employs yearly face-to-face target negotiations, and Dutch courts also have budget negotiations regarding their productivity with the Council for the Judiciary. In France the national court evaluation is primarily driven by the resource allocation process, to the point of being totally integrated to it. Hungary also has a central judicial administration, run by the National Office for the Judiciary, that electronically monitors all incoming cases and compares them to the number of resolved cases. Quantitative elements are most actively considered in the resource allocation. For the Netherlands, even monthly reports are sent to the Council for the Judiciary.

Another method employed to account for expenses to the general public are the annual reports on the performance of the courts by the courts themselves or by an inspectorate (Italy, Hungary, France, The Netherlands). Sometimes these reports are specifically aimed at a specific entity, e.g. the National Office for the Judiciary in Hungary, or the Council for the Judiciary in the Netherlands. It publishes annual reports with performance numbers for the courts. The above is summarized in the Table below, where evaluation methods are connected with the purposes of evaluations, in accordance with the scheme of Foss-Hanssen⁷³ (see Table 1).

6.6. Reflection on performance measurement and management in judiciaries.

Considering the size and diversity within the EU, with (currently) 28 member states, it is quite problematic to measure court and judicial performance accurately in a comparable way, collecting reliable data that can be validly aggregated and compared. National court administrations use different definitions of performances and therefore, even if data have been registered accurately and reliably, they cannot be validly aggregated beyond the national level. From a data collection perspective, the CEPEJ judicial systems study depends on the voluntary cooperation of the member states, and the way the data are gathered are often not ideal. Also during our meeting with experts it was said that the collection of data according to the protocol is a cumbersome exercise.

The EU has limited competences in the national justice fields, but improving the effectiveness of national justice systems is a priority.⁷⁴ This also implies the norms of the European Charter for fundamental rights, for example if it comes to fair trial rights and access to justice concerning the implementation of EU law in the member states.

⁷³ FOSS HANSEN, (2005), p. 451.

⁷⁴ Justice Scoreboard 2015, p. 2. Justice Scoreboard 2016, p. 1.

Table 1: Evaluation in the context of court administration and the judiciary

Evaluation in the context of court administration and the judiciary				
Purpose of evaluation	Subject of evaluation	Evaluation method	Actors	Responsibles
Professional learning.	Professional knowledge (e.g. law, social map, technology) and skills (e.g. conducting hearings, writing judgments, peer review)	Qualitative, horizontal Peer review	Judicial professionals, judicial managers, HRM department	Judicial professionals, judicial managers, HRM professionals
Appointment, Promotion of judges	Professional Performance	Exams, peer review, personal evaluation, hierarchy	Judicial Professionals, Judicial managers, Judicial Council and/or Ministry of Justice and/or parliament	Judicial professionals, policy makers in court administration
Upholding Judicial values	Professional behaviour	Peer review, horizontal; supervision, hierarchy	Judicial professionals, judicial managers	Judicial professionals, Judicial managers
Consistency of case law, quality of judicial decisions	Case law	Legal analysis, on a case by case basis.	Higher courts, judicial professionals, Legal Academics, Lawyers	Judicial professionals, Legal Academics, Lawyers
Management, Organisation development	Functioning and performance of the organisation Service provision	Supervision; Combination of quantitative and qualitative performance monitoring; Horizontal and Hierarchical Complaints, court user satisfaction surveys	Professional management and Judicial managers, Judicial professionals, Court users	Judicial professionals and Judicial management
Enhancing efficiency and timeliness in case management	Cases flow and case management,	Registration and counting and calculations	Court director Judicial management	Judicial management, Councils for the judiciary, Ministries of justice, Judicial Professionals
Political control, policy development	Performance of professionals and courts	Supervision Quantitative, Hierarchical	Management and court administration services	Management and judicial professionals, policymakers

Clearly, the scoreboard has a political perspective, and therefore the quantitative approach seems justified. The subjects of the scoreboard concern subjects for justice policies – and those belong to the domain of politics, but they also touch on professional judicial matters. Often, to achieve policy aims, somehow some cooperation between the court administration and the legislators/politicians is necessary. Timeliness of court proceedings, for example, may depend on many factors such as the rules of procedure, case management capacity of judges, behavior (of representatives) of parties, availability of court rooms, flexibility of deployment of judges, adequate support staff in the courts, the functioning of automated case flow management systems, adequate number of judges and so on. And politicians cannot instruct judges on how to manage their cases, as this would be a violation of judicial independence, but they can take measures that enable judges to work more efficiently.

As a matter of course, judges can be asked to live up to their professional responsibilities. They may be asked, for example to take responsibility for necessary and inevitable innovations, for example with a view to ICT's or to problem solving justice. Management of judicial performance and quality, however, should be in the hands of the judiciary, and also the evaluation of judicial professional performance. In this regard, it is also of importance on what level (individual, division, court, jurisdiction, national) data are gathered and shared with the political domain. Here, it may be helpful to discern between issues that belong to the political domain, issues that belong to management and organisation of the judiciary, the professional judicial domain, and issues that need cooperation between the political domain (a ministry of justice) and the judiciary. Usually, policy measures affecting judicial work are prepared while also asking the judiciary for advice. For example, new legislation may invoke a (temporary) growth of the influx of cases. All the more so, if policies aim at changes in, for example, access to justice, legal aid or changes in the judicial map, court fees and so on.

Summing up, measuring performance of courts and judges on the scale of the EU may be highly problematic from a methodological perspective, because the accuracy of the data gathered may be questionable and the data from one country may be not comparable with the data from another country – as is currently the case with data used in the Justice scoreboard. A (limited) unification of court statistics on the EU scale, of judicial performance standards and measurement and registration methods is indicated. The dearth of information assembled by the CEPEJ for the evaluation of judicial systems, therefore maybe flawed, because the required registration and data gathering processes are not standardized. For accurate information for policy development, the EU needs not only uniform performance standards and definitions, but also standardized data gathering and registration methods. The current methods for gathering data for the Justice scoreboard are not adequate.

A caveat is also indicated here. We have no clue whatsoever in how far all the measurement efforts in the countries we studied actually has contributed to policy measures. Often changes in justice policies, come from elsewhere: terrorism, austerity measures, problems with organised crime or transnational crime, coordination issues between the prosecutions office, the police and the courts; problems with family law, the development of ICT's, European policies and so on. Then, the question becomes most relevant: why should measurements be shared on the national, or even at the EU level? Central governments used to be a threat to judicial independence and impartiality; in some countries in Europe that is true again, in other countries central government policies try to improve judicial functioning by fighting corruption in the courts. This, in itself, gives enough ground to only ask measurement results at the level of the EU that are a derivative of national and local measurements for purposes of

professional learning and organisation development – using qualitative methods. Such information, however cannot be compared quantitatively.

Even although on the continent of Europe, judges are generally seen as civil servants with a special status, political influence on the ways judges hear and decide cases is considered undesirable. Cutting back expenses for the courts, because they took decisions that contradict the parliamentary majority, is considered not done. Any pressure exerted by politicians on judges directly or indirectly, open or covered, to influence their decision-making is not allowed. Nonetheless, some coordination between managing judges in the judiciary, the national court administration and the ministry of justice (and parliament) is necessary. And this coordination should be informed with data that show the current condition of the functioning of courts and judges on the national level. This information may be based on quantitative measurements, and may also consist of qualitative information concerning issues and achievements of the judiciary.

Passing performance information on to a central level usually does not only aim to give management information to those at the controls of the national justice system; it is also a way to account for the functioning of the courts in the democratic process.⁷⁵ That may involve acting strategically to prevent too much political intervention in court administration, and protect judicial independence.⁷⁶ Of course, this may generate a debate about the obligations of courts and judges concerning transparency.⁷⁷ And it is a task of politics to make choices on legislation and budgeting for the courts and the judiciary anyway. The quality of the information delivered to the politicians is crucial for the quality of their decision-making, and for the ways professionals in organizations may accept and react to political decisions. Furthermore, national court administrators may want to inform politicians about the (expected) effects of new legislation on the number of cases filed at the courts. Therefore a limited amount of information is necessary for coordination between managing judges, national court administrators and the ministry of justice (and parliament).

Below, we try a first answer on what should be measured and how, and for what purposes on the EU level.

7. Performance management and evaluation of courts' and judicial performances.

As courts and judges in our study function in a national and European environment in different contexts of justice policies, organizations and management, we considered how the paradox between accountabilities and the demand of judicial independence may be solved. The solution, on the one hand is that politicians should be adequately informed about performances, with a view to development of effective policies in the justice field. On the other hand, court administration and court management should leave enough room for professional autonomy of judges and for independent and impartial judicial case work and decision making. The latter sets limits on the data that should be gathered and shared.

⁷⁵ For a theoretical approach to accountability in general, see: BOVENS (2006).

⁷⁶ LANGBROEK, PHILIP. (ED.) (2010). Quality management in courts and in the judicial organisations in 8 Council of Europe Member States: A qualitative inventory to hypothesise factors for success or failure. Council of Europe, European Commission for the Efficiency of Justice: CEPEJ (2010).

⁷⁷ VOERMANS, W., (2007). Judicial transparency furthering public accountability for new judiciaries. Utrecht Law Review. 3(1), pp.148–159. DOI: <http://doi.org/10.18352/ulr.42>.

Judges have their own domains of professional responsibilities; they fulfil their function also in a court organization, and this organization has to be administered and managed.⁷⁸ This means that, in order to make both the political accountability and the management and organization development work, some interaction is necessary between judges and management, both with a view to professional performance and with a view to the management of the court organization. Furthermore, some interaction is necessary between the court administration and the political domain (ministry of justice, parliament) about the functioning of the courts, both organizationally and professionally). This can be done in many different ways, and the performances of the courts and judges are a subject in this interaction.

Generating enough information about the functioning of the courts to inform policymakers at the EU-level involves a standardization of norms for productivity and timeliness, plus a standardization of registrations. Other performance measures are to be derived of quality norms that relate to the services delivered by the courts: for example court users' surveys and their follow up. Further norms should relate to selection and appointment and promotion of judges, to lifelong learning standards for judges, to innovation capacity, and to upholding judicial values of judicial independence and impartiality. Debates should take place about what will be reasonable norms to be set on productivity, timeliness, court services, innovation and judicial independence and impartiality.

For now, we think that the following indicators will not intrude into the judicial domain of case management and decision-making. They will deliver enough information for the European Commission to see how a national judiciary performs.

Quantitative information can be gathered on:

- Production, in case categories
- Productivity, in case categories
- Costs, in Euro's
- Staff, and Judges in full time equivalents
- Timeliness, in case categories.

Qualitative information can be gathered on:

- Impartiality and independence - what indicators?
- Professional ethics and attitude – what indicators?
- Efforts to enhance judicial quality (within the judicial domain);
- Education permanente – what indicators?;
- Professional standards – what indicators?
- Efforts to enhance service quality (e.g. court user satisfaction surveys, mirror meetings)
- Quality of the interactions between politics, court administration, court management and judges.

This information can be gathered and reported on in annual reports, both in terms of numbers and descriptions of developments. For the gathering of information, definitions, and operationalization of definitions have to be designed, and standards must be developed for the reliable and accurate registration of data. Regular checks are necessary on the reliability and accuracy of those data. In order to reach this more restricted information should be gathered in a methodologically more sophisticated way than currently is being done for the Justice

⁷⁸ MIRKO NOORDEGRAAF, Risky Business: How Professionals and Professional Fields (Must) Deal with Organizational Issues, *Organization Studies* 32(10) 1349–1371, 2011.

scoreboard. We understand this is a challenge for the European Commission, that can be mastered with enough support from knowledge centers like universities and national court administrations.

References

- Abma, T. M. Noordegraaf, (2003) Public managers amidst ambiguity. *Evaluation* Vol 9(3): 285–306
- Aikman, Alexander B. (1994) *Total Quality Management in the Courts, a handbook for Judicial Policy Makers and Administrators*, National Center for State Courts, Williamsburg.
- Bjørnholt, Bente, Flemming Larsen (2014), The politics of performance measurement: 'Evaluation use as mediator for politics', Vol. 20(4) 400–411.
- Bohni Nielsen, Steffen and Nicolaj Ejler (2008), Improving Performance? Exploring the Complementarities between Evaluation and Performance Management, *Evaluation* Vol 14(2): 171–192,
- CEPEJ (2016)12, Measuring the quality of justice, Document prepared by the CEPEJ-GT-QUAL on the basis of the preparatory work of Mr Fabio BARTOLOMEO (CEPEJ member, Italy), 7 December 2016.
- Contini, F. and Mohr, R. (2008). *Judicial Evaluation, Traditions, Innovations and Proposals for Measuring the Quality of Court Performance*. Saarbrücken: VDM-Verlag
- Dahler-Larsen, Peter (2000), Surviving the Routinization of Evaluation, *The Administrative Use of Evaluations in Danish Municipalities, Administration & Society*, vol. 32, 1: pp. 70-92,.
- Dori, Adriani (2015) the EU Justice Scoreboard –Judicial Evaluation as a New Governance Tool, Max Planck Institute Luxembourg Working Paper 2, pp. 1-35.
- Dunleavy, Patrick, Helen Margetts, Simon Bastow and Jane Tinkler (2005), "New Public Management Is Dead—Long Live Digital-Era Governance", *Journal of Public Administration Research and Theory* vol. 16, pp. 467-494
- Fabri, Marco and Philip M. Langbroek, (Eds.) (2000), *The challenge of change for judicial systems in Europe. Developing a public administration perspective* Amsterdam/Oxford/Washington 2000: IOS Press
- Garbarino, Sabine, Jeremy Holland (2009), *Quantitative and Qualitative Methods in Impact Evaluation and Measuring Results* Issues Paper, GSDRC Emerging Issues Research Service UK, <http://www.gsdrc.org/docs/open/eirs4.pdf>
- Hagsgård, Marie B. (2008). *Internal and External Dialogue: A Method for Quality Court Management*. *International Journal for Court Administration*. October 2008 Issue. Available at http://www.iaca.ws/files/LWB-Marie_Hagsgard.pdf [Accessed 24/06/2012]
- Hagsgård, Marie B., 2014. *Internal and External Dialogue: a Swedish Approach to Quality Work in Courts*. *Oñati Socio-legal Series* [online], 4 (5), 993-1014. Available from: <http://ssrn.com/abstract=2541449>

Foss Hanssen, Hanne (2005) Choosing Evaluation Models, A Discussion on Evaluation Design, *Evaluation*, Vol 11(4): pp. 447–462

Hoerner, Julian and Paul Stephenson (2012), Theoretical Perspectives on Approaches to Policy Evaluation in the EU: the Case of Cohesion Policy, *Public Administration* Vol. 90, No. 3, (699–715).

Johnsen, John T. (2012). “The European Commission for the efficiency of justice (CEPEJ). Reforming European justice systems – ‘Mission impossible?’”, *International Journal for Court Administration*, Special issue, 1–19. Available at: http://www.iaca.ws/files/ijca_2012_special_issue/ijca-2012-special_issue.pdf

Joyce, Philip G., Susan Sieg (2000), Using Performance Information for Budgeting: Clarifying the Framework and Investigating Recent State Experience. Paper Prepared for the 2000 Symposium of the Center for Accountability and Performance of the American Society for Public Administration, held at the George Washington University, Washington, D.C., February 11 and 12,

Keilitz, Ingo, Daniel Hall, Laurence Glanfield, (2017) *Global Measures of Court Performance*, Second Edition, 2017

Langbroek, Philip (2010) (ed) and research director, *Quality management in courts and in the judicial organisations in 8 Council of Europe member states*, CEPEJ 3

Lewis, Jenny M. and Peter Triantafillou (2012) From performance measurement to learning: a new source of government overload? *International Review of Administrative Sciences* 78(4) 597–614,

Mayne, John (2007). Challenges and Lessons in Implementing Results-Based Management, *Evaluation*, issue 13; p.87-109

Miedema, Jitta, Abeltje Hoogenkamp, Cora Van Steenderen Koornneef, Maria Mul (2015), *Spiegelbijekomsten, de stand van zaken na 5 jaar, Feedback vanuit de relevante buitenwereld*, *Tijdschrift voor de Rechterlijke Macht*, p.315-320.

Ng, Gar Yein, Marco Velicogna and Cristina Dallara (2008), *Monitoring and Evaluation of Court System: A Comparative Study*, CEPEJ Strasbourg.

Noordegraaf, Mirko (2013), Reconfiguring Professional Work, Changing Forms of Professionalism in Public Services, *Administration & Society*, vol. 48, 7: pp. 783-810.

Noordegraaf, Mirko (2011), Risky Business: How Professionals and Professional Fields (Must) Deal with Organizational Issues, *Organization Studies* 32(10) pp.1349–1371.

Osborne, David and Gaebler, T. (1993) *Reinventing Government, how the entrepreneurial spirit is transforming the public sector*, Boston MA, Addison Wesley.

Patton, Michael Quinn (2014): *Qualitative Inquiry Inutilization-Focused Evaluation*, in: Goodyear, Leslie, Barela, Eric, Jewiss, Jennifer *Research Methods for the Social Sciences: Qualitative Inquiry in Evaluation : From Theory to Practice*, Jossey-Bass, San Francisco, pp. 25 -54

Peršak, Nina (2014), Shared standards of justice: Towards normative indicators of judicial legitimacy at EU level, in: Lieven Pauwels, Gert Vermeulen (eds), *Actuele ontwikkelingen inzake EU-justitiebeleid, cannabisbeleid, misdad en straf, jongeren en jeugdzorg, internationale vrede, veiligheid en gerechtigheid, gewelddadig extremisme & private veiligheid en zelfregulering*, Maklu Antwerp, pp. 45-65.

Piana, Daniela (2016), Reforming the judiciary through standards: agency empowerment and centre (re)building in Italy, 2001–2015 *International Review of Administrative Sciences* *International Review of Administrative Sciences*, 0(0) pp 1–17.

Pollitt, Christopher, (2013) The logics of performance management, *Evaluation*, Vol 19(4) pp.346–363.

Schillemans, Thomas, (2011) Does Horizontal Accountability Work? Evaluating Potential Remedies for the Accountability Deficit of Agencies, *Administration & Society*, vol. 43, 4: pp. 387-416.

Shapiro, Martin (1981), *Courts, a comparative and political analysis*, University of Chicago Press.

Soltani, Ebrahim, Robert van der Meer, John Gennard, Terry Williams (2003), A TQM Approach to HR Performance Evaluation Criteria, *European Management Journal* Vol. 21, No. 3, pp. 323–337.

Steckler, McLeroy, Goodman, Bird McCormick (1992) Towards Integrating Qualitative and Quantitative Methods: An introduction *Health Educ Q*. Spring;19(1): pp.1-8.

Volker, Mathias, Michael Kalde and Ralf-Günther Lüpkes, (2008) Qualitätsmanagement der Oberlandsegerichte. *Deutsche Richterzeitung*, Oktober, pp. 269-271.

Voermans, W., (2007). Judicial transparency furthering public accountability for new judiciaries. *Utrecht Law Review*. 3(1), pp.148–159. DOI: <http://doi.org/10.18352/ulr.42>.

Vosselman, Ed (2013) Accounting, Accountability, and Ethics in Public Sector Organizations Toward a Duality Between Instrumental Accountability and Relational Response-Ability *Administration & Society*, vol. 48, 5: pp. 602-627.

Vries, Michiel S. de (2010) Performance measurement and the search for best practices, *International Review of Administrative Sciences*, vol. 76, 2: pp. 313-330.