Oñati International Institute for the Sociology of Law
Instituto Internacional de Sociología Jurídica de Oñati
Oñati Socio-Legal Series
ISSN: 2079-5971

Oñati Socio-legal Series, v. 4, n. 5 (2014) – Evaluating Judicial Performance
ISSN: 2079-5971

Formula over Function? From Algorithms to Values in Judicial Evaluation

FRANCESCO CONTINI*
RICHARD MOHR*
MARCO VELICOGNA*

Contini, F., Mohr, R., Velicogna, M., 2014. Formula over Function? From Algorithms to Values in Judicial Evaluation. *Oñati Socio-legal Series* [online], 4 (5), 1099-1116. Available from: http://ssrn.com/abstract=2533902

Abstract

This paper discusses the forms and effects of the 'invasion' of the 'temples of the law' by new economic and managerial forms of performance evaluation. While traditional judicial evaluation focused on how to select and promote individual judges and on the legal quality of the single case, new quantitative methods and formulas are being introduced to assess efficiency, productivity and timeliness of judges and courts. Building on two case studies, from Spain and the Netherlands, the paper illustrates two contrasting approaches to judicial performance evaluation. On the one hand individual judges' productivity is evaluated through quantitative data and mathematical algorithms: in the extreme case considered here, judge's

Article resulting from the paper presented at the workshop *Evaluating Judicial Performance* held in the International Institute for the Sociology of Law, Oñati, Spain, 9-10 May 2013, and coordinated by Francesco Contini (National Research Council of Italy), Jennifer Elek (National Center for State Courts), Kathy Mack (Flinders University), Sharyn Roach Anleu (Flinders University) and David Rottman (National Center for State Courts).

* Francesco Contini is a senior researcher at the Research Institute on Judicial System of the Italian National Research Council (IRSIG-CNR), where he is coordinating the research area "Quality assessment in justice systems". He studies the institutional transformations of European judiciaries with focuses on case management, performance evaluation and e-justice. He has written various articles and books on such topics, among which "Judicial Evaluation" with Richard Mohr (VDM 2008), and 'ICT and innovation in the public sector', edited with Giovan Francesco Lanzara (Palgrave 2009). He has collaborated with international organizations in judicial reform programmes in Europe, Africa, and Asia. Istituto di Ricerca sui Sistemi Giudiziari (IRSIG-CNR), Via Zamboni, 27 Bologna. francesco.contini@irsig.cnr.it

* Richard Mohr is director of Social Research, Policy and Planning, is a visiting professorial fellow in the Legal Intersections Research Centre at the University of Wollongong, Australia, and maintains connections with the Institute for Judicial Research in Bologna (IRSIG-CNR), publishing with Francesco Contini. He has published in socio-legal and semiotic journals. His most recent book is *Law and Religion in Public Life: The Contemporary Debate*, edited with Nadirsyah Hosen (Routledge, 2nd, paperback ed. 2013). He has written and researched in many aspects of court services and evaluation, and on technological limits to legal performativity. Social Research Policy and Planning PL, P.O. Box 338, Thirroul, NSW 2515, Australia. rmohr@srpp.com.au

* Marco Velicogna is a researcher at IRSIG-CNR, where he studies the functioning and change of judicial systems focusing on e-Justice, management, evaluation and quality. He has participated in several national and international research projects including a key role in the EU co-funded Large Scale Pilot project e-CODEX. He has also served as consultant for the Italian Ministry of Justice and collaborated with several international institutions such as UNODC, DG-Justice (EU Commission) and OSCE, and as scientific expert for the European Commission on the Efficiency of Justice of the Council of Europe (CEPEJ-CoE). Between other publications, he is author of CEPEJ Study No. 7, 'Use of Information and Communication Technologies (ICT) in European Judicial Systems' (2008). Istituto di Ricerca sui Sistemi Giudiziari (IRSIG-CNR), Via Zamboni, 27 Bologna. marco.velicogna@irsig.cnr.it

remuneration was adjusted accordingly. On the other hand quantitative and qualitative data, collected by a variety of methods and theoretical frameworks, are used as the basis of a multi-layered negotiation process designed to find a synthesis between competing economic, legal and social values aimed at improving overall organizational performance. Considering the flaws of unidimensional measurement and evaluation systems and considering the incommensurability of the results of the multiple evaluative frameworks (economic, legal, sociological) required to overcome such flaws, the authors argue there is a need for political dialogue between relevant players in order to allocate the values appropriate to judicial evaluation.

Key words

Judicial evaluation; performance studies; management; Spanish Judiciary; Dutch Judiciary

Resumen

Este artículo analiza las formas y efectos de la "invasión" de los "templos de la ley" por nuevas formas económicas y de gestión como la evaluación del rendimiento. Mientras que la evaluación judicial tradicional se ha centrado en la forma de seleccionar y promocionar a jueces individuales, y en la calidad jurídica de un caso individual, hoy en día se están introduciendo nuevos métodos cuantitativos y fórmulas para determinar la eficiencia, productividad y oportunidad de jueces y tribunales. A partir de dos estudios de caso de España y los Países Bajos, el artículo ilustra dos enfoques opuestos de la evaluación del rendimiento judicial. Por un lado la productividad de jueces individuales se evalúa a través de datos cuantitativos y algoritmos matemáticos: en el caso extremo que se considera aquí, la remuneración del juez se ajustó en base a la evaluación realizada. Por otro lado, se utilizan datos cuantitativos y cualitativos, recogidos mediante diversos métodos y marcos teóricos, como base de un proceso de negociación en múltiples niveles, diseñado para encontrar una síntesis entre valores económicos, legales y sociales, destinados a mejorar el rendimiento general de la organización. Teniendo en cuenta los defectos de los sistemas de medición y evaluación unidimensionales, y considerando la inconmensurabilidad de los resultados de los marcos de evaluación múltiples (económicos, jurídicos, sociológicos) que se requieren para superar esos defectos, los autores sostienen que hay una necesidad de diálogo político entre los actores implicados para asignar valores adecuados a la evaluación judicial.

Palabras clave

Evaluación judicial; estudios de rendimiento; gestión; Poder Judicial Español; Poder Judicial Holandés

Table of contents

1. Introduction

This paper discusses the relationships between individual and organisational performance evaluation in civil law justice systems and appraises the contribution of quantitative methods in each case. These issues are discussed in the context of the social, organisational and technological changes affecting judicial systems. It takes two case studies, from Spain and the Netherlands, which illustrate ways of approaching evaluations of individual judges and of judicial organisations, respectively.

Until a few years ago, judicial evaluation was considered mainly as a matter of how to select and promote individual judges. This was consistent with the court as a 'temple of the law' in which the judge works as an isolated decision maker informed purely by traditional legal values such as fairness and impartiality, and protected from improper influences by a set of constitutional principles and institutional arrangements, in particular institutional and individual independence. The staff supporting the judge were simply providing ancillary services, useful but largely independent of the content and the organisation of judges' work. In this setting, the judges' contribution and performance were largely independent of organisational constraints and support. Individual evaluation with a focus on traditional legal values was considered one of the most important parts of the administration of justice.

A number of factors changed this picture. Budget cuts and blowouts in litigation, the introduction of managerial principles and of pervasive information and communication technologies (ICTs), active case management and the involvement of stakeholders in the organisation of judicial proceedings have all contributed to making the judge and the court organisation much more integrated, service oriented and efficiency driven.

The judges' task is now supported by a growing set of tools; (quasi) judicial activities are delegated to non-judicial staff. Case management and document drafting, case filing and summoning are increasingly delegated to (and inscribed into) large technological systems. Resources are allocated considering new parameters, including the expected caseload and the actual productivity of courts. And most important, the judge is not only a passive user of this new set of organisational tools. Judges are expected to promote and use these new arrangements. The judge cannot be evaluated just as an independent decision maker, but also as an organisational actor. Accordingly new parameters and new evaluative methods have been introduced, and also the entire organisation has to be evaluated. Indeed, the service that the court provides to its users is the result of a collective effort, and not just of the contribution of the final decision maker. Accordingly organisational evaluation introduces a different unit of analysis (the organisation), different evaluative methods and new public values (e.g. productivity and efficiency).

The paper discusses how these new evaluative approaches work at individual and organisational levels analysing two judicial evaluation mechanisms introduced in the last ten years in European judiciaries.

The first approach (*módulos de dedication*) was developed by the Spanish Judicial Council to introduce a system of performance-based remuneration for judges. Section 1 describes its ratio and functioning, and Section 2 makes an assessment of its impact on individual judges' behaviour.

The second approach, taking an organisational focus, was developed by the Dutch judiciary to finance courts considering their expected workload (called *Lamicie*). Sections 3 and 4 consider the context in which this performance based budgeting has been developed and how it pushes courts to increase efficiency. Furthermore, within the Dutch system, the Lamicie component is balanced by a second component (*RechtspraaQ*) developed to improve the quality of organisational

processes from a multidimensional perspective. This balances the risks of courts focussing purely on narrow measures of efficiency, to the neglect of other important values (Section 5). The concluding section discusses the limits of unidimensional evaluation systems and highlights the questions of incommensurability affecting the results provided by evaluative frameworks based on different approaches (economic, legal, sociological) designed to overcome the flaws of unidimensional evaluations. This finding leads to the intrinsic political nature of judicial evaluation, and therefore to the need of political dialogue between relevant players to find viable syntheses to multidimensional evaluation endeavours.

2. Individuals: the performance-based salary of the Spanish judges

The genesis of the output measurement system (*módulos de dedicación*) of the Spanish judiciary dates back to the end of the eighties. It was established by the Judicial Council (Consejo General del Poder Judicial) to determine the number of judges and other personnel required per court, based on output.

In 2000, after more than ten years and criticism of the lack of rigour in measurement (Signifredi 2006, Contini and Mohr 2008, p. 40)[1], the *Consejo* approved a new set of *módulos*, intended to measure the workload of judges at the individual level. The new system, based on a more accurate statistical method and data collection than earlier versions, assigns a standard time (or value) to each type of proceeding, calculated considering the degree of difficulty and the time it requires (Consejo General del Poder Judicial 1999, p. 3). *Módulos* are calculated taking into account bench time and other case related tasks such as preparing and studying the case, and meeting with parties. They distinguish between cases decided by a single judge and by panels. The result of this analysis is a matrix which gives a standard time required to handle each major type of proceeding, as shown in table 1 (Poblet and Casanovas 2005, Decker *et al.*, 2011, p. 95)[2]. The system also estimates non case-related times, such as those consumed between hearings, in breaks, for administration, training etc. that is subtracted from the total working time, to derive a measure of the time to be dedicated to case related activities.

In this way the *módulos* establish the number of case-related working hours per year and per each type of judicial position. The number of case-related working hours becomes the benchmark against which the productivity of each judge is to be evaluated. At the end of each semester, the evaluation of the performance of each individual judge can be carried out comparing the actual and the expected workload. The algorithm is:

> (number of cases decided per each type of case x standard working hours of each type of case) / number of case-related working hours per year.

The table below provides an example of *módulos* of the First Instance Civil Court. The productivity of each judge will be determined calculating the value, established by the table, for each case decided during the year.

[1] The measures did not take into account weightings for different types of cases.

[2] A special numerical value is provided for exceptional cases requiring an inordinate amount of a judge's time on a case-by-case basis.

Table 1: Civil Courts of First Instance

CIVIL COURTS OF FIRST INSTANCE

White Paper Module (based on cases docketed)	850 civil actions
WORKLOAD MODULES (based on the number of cases concluded)	
Civil Actions Concluded	Points
Proceedings for Large Claims	12
Proceedings for Lesser Claims	3.25
Delaratory Proceedings	2
Oral Proceedings	1.4
Summary Executory proceedings	1.25
Mortgage Proceedings	1
Eviction Proceedings	1.25
Bankruptcy Proceedings	12
Other Civil Actions	1
Cases Requiring an Exceptional Amount of Time	up to 130

Total annual hours	1,650 hours
Time subtracted to provide for activities not reflected in the table above (20% of the annual number of work hours, or -7.5 hours per week)	330 hours
Module for Civil Courts of First Instance	**1,320 hours/points**

Source: (Consejo General del Poder Judicial 1999, p. 10).

In 2003, the law n. 15 establishing the new remuneration system for judges and prosecutors stated that salary was to be composed of a fixed and a variable component, based on individual performance (art. 2.1). This variable component was to be based on 'transparent criteria' established by regulation (art. 2.3). The *módulos* were identified as the tools for this purpose and, since 2004, connected individual judges' performance and salary (Bagues and Esteve-Volart 2010)[3]. The new salary plan provided a fixed amount (base salary) plus a variable amount (based on productivity). Judges who exceeded the productivity standard by more than 20% would receive an additional remuneration of 5% of their salary. The base salary of judges who performed less than 80% of the applicable *módulos* could be cut by 5% (Decker *et al*. 2011, p. 97-98).

The introduction of this remuneration system drew strong criticism from the Spanish judges. The Judicial Council never used the system to reduce the salary of the less productive judges, but only to reward the most productive (Contini and Mohr 2007). In 2006 the Spanish Supreme Court ruled that the *módulos* system violated the principle of financial independence of the judiciary and that the variable remuneration scheme was not based on sufficiently objective, equitable and transparent principles (Decker *et al*. 2011, p. 97-98). After further consultation between the Judicial Council and the judges' professional associations, the system was reintroduced in 2008 with some changes in the remuneration schema[4]. According to the new rules, 'judges whose performance exceeded the benchmark by 20% would still be rewarded with a 5% increase in their salary, but now judges who were merely complying with the benchmark would also qualify for a bonus. In particular, judges whose production was between 100% and 120% of the benchmark were to receive a 3% bonus. In contrast, judges whose production did not reach 80% of the benchmark would not be penalized in any way' (Bagues and Esteve-Volart 2010, p. 2).

[3] Agreement of December 3, 2003 of the Consejo General del Poder Judicial, approving the regulation n. 2/2003, implementing the law 15/2003, de 26.

[4] As a consequence of the *Acuerdos del Pleno del CGPJ de 19.12.2007* and of the *Acuerdos de la Comisión Permanente del CGPJ de 11/12/2007*.

2.1. Impact of performance based remuneration

Bagues and Esteve-Volart argue that introduction of the first performance based remuneration scheme 'resulted in an average increase in production, though the effects were mixed across the distribution' (Bagues and Esteve-Volart 2010, p. 2). Comparing first semester 2003 data (pre-performance-based evaluation) with second semester 2005 data (when the evaluation was in place), they observed that the remuneration scheme improved overall production by 7%, with a corresponding rise in costs of 2%. There was an increase in the number of judges producing above 120%, mainly driven by an increase in the number of judges producing just above 120% and so getting the salary bonus. In the same period, the number of judges producing more than 160% decreased by approximately 25%. It is also possible to observe a small production peak around 100%, which the authors suggest may indicate 'that some judges have decided to adjust their production to level of production which is expected from them by the new system' (Bagues and Esteve-Volart 2010, p. 6).

The second performance based remuneration scheme, introduced in 2008 was analysed using first semester 2008 data. In overall production, compared to 2003 data, the second system resulted in an increased production of 7% with a corresponding rise of 2.6% of the costs. A comparison between the 2008 data with those of 2005 shows a slight production decrease and a cost increase of 0.7%. More interesting, from 2005 to 2008 there is a reduction in the number of judges producing just above 120% (5% bonus) and an increase of judges producing just above 100% (3% bonus). At the same time, the removal of the 80% threshold resulted in an increase of judges producing less than 80% (Bagues and Esteve-Volart 2010, p. 7). Incentives make a difference.

Remuneration schemes have an impact on the judges' productivity, and judges can adjust their output to meet the expectations embedded into the performance based salary schema. As rational actors, a certain number of judges seem to enact strategic behaviours adjusting their output to maximise their benefits based on the incentive schema. This is not an ethical or moral judgment, but a reasonable inference from the data and the study we are discussing. There is the rule of law, but there is also the rule of economics. Strategic behaviour oriented to minimise the effort required to receive incentive payments is simply a rational, goal oriented choice. Introducing extrinsic incentives also has the effect of lowering the intrinsic motivation of judges as the reduction of the number of judges producing more than 160% shows (Bagues and Esteve-Volart 2010, p. 8). This latter phenomenon cannot be attributed to rational economic behaviour. There is no reward for producing less. It might be inferred, however, that when workload is reduced to a scheme of incentive payments, other rewards are diminished. These may have been pride in work or honour derived from high levels of productivity, whose currency was devalued by hard cash.

To sum up, individual productivity is the primary and unique evaluative criterion considered by the Spanish *módulos*; individual incentives have some impact and judges respond to incentives, increasing their output, with a concentration of judges just over critical thresholds; there is also evidence that judges who had been considerably more productive than the criteria required have reduced their output, and that the abolition of the possibility of salary reduction has increased the number of judges producing below this threshold.

It can be questioned whether the mixed (or negative) results of the remuneration schema in terms of productivity are due to an inappropriate incentive system. In theory, salary reduction for less productive judges could be reintroduced (as in the 2004 model) as well as additional remuneration for judges producing more than a given threshold (such as 160%) to maintain consistent incentives at all levels of productivity. This may have positive effects on individual productivity and reward super-productive judges. However, a unidimensional measurement system will only

have impacts on that one dimension. Thresholds for incentives bring productivity both up and down to meet that threshold, and the introduction of financial incentives may block or devalue the perception of other values. As noticed, If salaries are adjusted based on productivity alone then other values, such as the legal quality or clarity of the decisions, may be compromised (Domenech Pascual 2008). Judicial performance should not be confused with productivity. Other values and other criteria have to be applied, and included in the judicial evaluation mechanism.

Some Spanish judges seem to agree with this consideration, and in the last few years various cases have been filed at the *Audiencia National* (Supreme Court) to challenge the *módulos* or the remuneration system. In one of such cases, filed by the *Foro Judicial Independiente* (one of the Spanish judges' associations), the Supreme Court ruled against the *módulos* and the annexed salary based performance. In particular in February 2012 the *Sala tercera* of the Court ruled against (Tribunal Supremo 2012) the 'agreement' between the Judicial council and the Judicial Association that approved the ranking of judges based on their individual performance drafted by the Judicial Inspectorate. One of the reasons for the ruling is the agreement's lack of any appropriate legal basis. But more interestingly, the Court stated that, 'The object truly relevant for the purposes of the present appeal is not so much the approval of the lists made by the Inspectorate [stating the rank of judges based on individual productivity], but the approval of a general criterion upon which the listing was made'. Indeed, the *módulos* do not take into account 'all the tasks to be undertaken by the judge or magistrate, nor all relevant aspects of their activity in court, but only the most easily observable and measurable. They bypass important aspects of performance that are not assessed and, therefore, do not count for compensation: the quality of judicial decisions taken, their agreement or disagreement with the legal system, as well as, among others, the activities of judges and magistrates who are not formalized in a jurisdictional decision'.

This leaves the question of how to translate the legal provision of having transparent criteria measuring the individual performance of judges established by the law into an accurate, appropriate and not misleading performance measurement system. Is it a problem that can be faced with better injection of economical, statistical or mathematical science or does some other discipline need to be brought on board? More profoundly is it realistic and effective to assume that performance can be measured through clear, transparent and objective criteria and methods? Or do other assumptions and methods have to be considered?

The next case study shows a possible solution for these problems, through an organisational rather than individual focus, a wider range of responses, beyond individual salary incentives, and a wider range of criteria, methods and forms.

3. Organisations: the performance-based budget of the Dutch courts

In the Netherlands, as in Spain, workload measures are one of the pillars of the evaluation mechanism. The Dutch Judicial Council has developed a system, called *Lamicie*, to measure the average number of minutes required to handle each different type of case and define resource allocation according to the expected production.

The *Lamicie* model was developed as part of a re-organization of the Dutch Justice Administration. In 2002 two relevant norms were introduced in the Netherlands, resulting in the reshaping of the structure of the Judicial Administration. Firstly, as a result of the Act on the Council for the Judiciary (*Wet Raad voor de rechtspraak*), courts (district and appeal) are supervised by a 'Council for the judiciary, to which they are financially and organisationally accountable' (Langbroek 2010, p. 86-87). The newly established Council assumed responsibility for a number of duties previously carried out by the Minister of Justice (Council for the Judiciary 2008).

'The Council for the Judiciary allocates funds to the courts in accordance with their production, but also enhances the organization development of the courts and the judicial organization as a whole' (Langbroek 2010, p. 86-87). Secondly, the Dutch Judiciary Organisation and Management Act (*Wet organisatie en bestuur gerechten*) established, in each court, a management board consisting of the court manager, the court president and the presidents of the court sectors. To balance the increased independence of the court system with a proportionate level of accountability these organizational changes were linked to the introduction of a new workload based accounting and financing system, the *Lamicie* model. 'The system is intended as a tool of accountability for the court as a whole and not for the evaluation of the performance of the individual judge' (UNODC 2011, p. 120). This marks a clear difference from the Spanish system that has been developed as a tool to monitor, rank and increase the productivity of the individual judge. Furthermore, the Dutch 'approach and the workload measures have been developed within the judiciary and with the support of the judges, resulting in relatively little criticism' (UNODC 2011, p. 120).

Between 2002 and 2005 the *Lamicie* accounting system was introduced into the courts. The Court Sector (Funding) Decree of 2005 completed the system introducing output based funding to judicial administration, bringing about a shift in financing 'from a cash commitments system to a cost-benefit system' (Council for the Judiciary 2013, p. 2). According to the new financing system each court and the judiciary as a whole 'receive money in accordance to the production of cases in the year previous to the budget year' (Langbroek 2010, p. 86). A small part of the budget, however, is a lump sum allocation for special projects and special costs of court-proceedings (e.g. hiring experts by the courts)'.(Langbroek 2010, p. 87)

The core of the system is the *Lamicie* workload measurement system. In the same spirit as the Spanish *módulos*, the *Lamicie* model differentiates between 53 different case categories. For each of these categories the model provides an estimation of the average court time (in minutes) that is needed to handle a case (i.e. judge and court staff time needed to prepare and finalize case). So, for example, according to the *Lamicie* model (2005), a defended employment dismissal case requires 303 minutes of judge time and 205 minutes of staff time for a first instance court while, for the same court, a rent case requires 110 minutes of judge time and 161 minutes of staff time (van der Torre *et al*. 2007). The number of minutes estimated for each case category 'is based on workload measurement to be repeated every 3 years' (Langbroek 2010, p. 86-87). The time allocation surveys are conducted in the courts by external researchers. As with the Spanish *módulos*, *Lamicie* also establishes the 'yearly expected number of working hours per each judge and staff unit (e.g. 1,137 hours in 2003, 1,151 hours in 2004)' (UNODC 2011, p. 120).

However, unlike the *módulos*, *Lamicie* and its complex mathematical and statistical system is used to define the standard time and cost per case as a basis of the negotiation of the allocation of financial resources through a two tiered system, from the Ministry of justice to the Council and from the Council to the courts (Council for the Judiciary 2013, p. 3). It is not used to assign financial incentives to the most productive judges. In simplified terms[5], the resources that the judiciary and the courts are entitled to receive is defined on the basis of the following algorithm:

(number of cases per case category) x (minutes per case) x (price per minute).

[5] The system is more sophisticated but the discussion of such details is irrelevant for the purposes of this work.

3.1. The Lamicie model in practice

At a higher level, the allocation of financial resources to the judiciary is decreed by law on an annual basis (Council for the Judiciary 2004). The 'Council for the judiciary receives money from the Ministry of Justice (as a part of the budget bill of the Ministry) according to the aggregate production of all the courts together' (Langbroek 2010, p. 86-87). About 95% of the annual contribution from the Ministry to the Council 'is attributable to output funding (price x quantity [of case disposals])' (Council for the Judiciary 2013, p. 3). Council and the Ministry negotiate on the basis of the price of 10 categories of cases, which encompass the 53 categories used at court level. Prices are negotiated for a three-year term on the basis of the results of the time allocation surveys. So for example, for the period 2011-2013, prices vary 'from €140 for a sub-district court case to €3,615 for a civil case before a court of appeal' (Council for the Judiciary 2013, p. 4).

Each year the Council submits to the Ministry of Justice a proposal for the number of case disposals. The Ministry uses it as a basis for the Ministry budget to be submitted to the Parliament indicating the number of cases it proposes to fund. If the number of court cases differs from the number proposed by the Council, the Ministry has to explain the difference (Council for the Judiciary 2013, p. 3). The Council is then accountable to the Minister of Justice for the way in which these resources are spent by the judiciary.

In its turn, the Judicial Council allocates funds to the courts mainly considering their production as established by the *Lamicie* model [6] (Langbroek 2010, p. 86-87).

At court level, financial allocation is based on a planning and reporting cycle with yearly plans, progress reports (every four months) and annual reports involving the Judicial Council and the Management board. The board prepares 'an annual plan of activity which includes productivity targets and a budget proposal based on the *Lamicie* model' (UNODC 2011, p. 120). Then plans are submitted to the Council for the Judiciary, negotiated and, finally, agreed. The budget itself is calculated on the basis of the number of cases the court is expecting to handle and the price of the cases. The price of cases is fixed annually by the Council for each of the 53 categories taking into account the results of the time allocation surveys, though it has been noted that in practice their influence is limited (Council for the Judiciary 2013).

At court level, this workload-based budgeting mechanism triggers discussions about how the court is operating, how to use the available resources, and how to improve organisation and procedures. Indeed, once the budget is approved, the board is free to decide how to use the available resources (Council for the Judiciary 2004, p. 11, UNODC 2011, p. 120). Various options can be considered such as organisational and case management improvements (technologies, protocols with case parties, mediation etc.), delegation of quasi-judicial activities to specialised law clerks, employment of part time judges etc. In any case, this budget allocation mechanism provides incentives to improve court organisation and look for better ways to deploy resources.

This involves the entire court. Indeed, even though the decision about the budget planning and allocation is formally entrusted to the management board, the decision benefits from discussions involving judges and staff at different levels[7]. Once the annual cycle of activities has been completed, each court submits an annual report to the Council on how far the plan has actually been realized. This report provides another venue of performance evaluation: it includes a calculation

[6] A small part of the budget, however, is allocated lump sum, in relation to special projects and special costs for court-proceedings such as hiring experts by the courts.

[7] Interviews with court presidents and ex-court presidents carried out in 2012 and 2013 by one of the authors.

and accountability for the financial administration of the court for the previous budget year, the way in which the necessary activities taken from the budget were carried out and an audit of spending. The Council then evaluates how well the courts have met the goals based on the previous year's plan. If the court's productivity has been up to 5% greater than the established target, the court funding for the following year is increased by 70% of the value of the overproduction. If the productivity has been lower than the target, other measures may be taken (UNODC 2011, p. 120). No doubt, *Lamicie* has injected a massive dose of statistics, mathematics and economics into court operations. But the results of the measurement systems, as noted, are not transformed automatically into consequences. Rather, there is room for discussion and negotiation at various levels: Ministry of Justice and Judicial Council, Judicial Council and each court, and within the court. Here, the evaluation provided by *Lamicie* is translated into organisational measures addressed to improving court operations. However, it was clear also to the Dutch judiciary that *Lamicie* would have provided a strong push toward efficiency and productivity to the possible detriment of other values. Therefore, from the beginning the pressure toward efficiency was balanced by a complex quality management system called *RechtspraaQ*.

3.2. Balancing quantity and quality: the RechtspraaQ

The quality management system was initially provided for by the Judicial Organisation Act and then, in 2005, by the Order in Council on the Financing of the Administration of Justice.[8] It provides a framework for the evaluation and promotion of judicial quality and performance that is much broader than the one embedded in *Lamicie*. The term 'quality' in the judicial context usually refers to the legal (substantive) quality of court judgments. This form of quality is traditionally monitored through the system of appeals. More and more, however, it is becoming 'accepted that the quality of the judicial system also includes aspects connected with the service provided by the courts. Examples include the lead times and the manner in which customers (litigants, counsel, etc.) are treated. The quality system applied by the courts promotes the quality of the judicial system in the broad sense' (Council for the Judiciary 2008, p. 3). Overall, 'the quality system *RechtspraaQ* deals mostly with organisational quality rather than juridical (although issues of integrity, independence and legal unity are addressed)' (Ng 2007, p. 84).

The *RechtspraaQ* is composed of quality regulations, measuring systems and other components. The normative function consists of quality regulations of the courts and the judicial performance measuring system; the measuring function consists of court-wide position studies, client evaluation surveys, staff satisfaction surveys, peer reviews and audits. The other elements are the complaints procedure and peer review.

Each court and each court sector has to define its own quality regulation, based on models developed by the Council for the Judiciary in cooperation with the courts (Langbroek 2010, p. 85), which can be adapted to the specific circumstances of each court. The regulations contain a structured description of the court's quality targets, but not the means by which they are to be achieved, since the courts are free to decide how they should operate (Council for the Judiciary 2008).

RechtspraaQ entails indicators for work processes, personnel policy, personnel and customer research, and also for the judicial function. It also provides indicators on the independence and integrity of judges, expertise, uniformity of the law, comportment (treatment of litigants and defendants), speed and timeliness. It is therefore a completely different system, designed to balance the drive towards productivity and efficiency provided by *Lamicie*. In sum, the *RechtspraaQ* aims at

[8] *Besluit financiering rechtspraak* 2005, Stb. 2005, nr. 55 (Order in Council on the Financing of the Administration of Justice).

evaluating and improving the quality of justice, both in its organizational component and in its final product (court decisions) (Ng 2007, p. 84).

There are several measuring instruments used within *RechtspraaQ*. The *assessment of the position of the court* is a tool to analyse, every two years, the progress of the improvement activities that are carried out in specific 'areas of attention' established by the quality regulations and to create new improvement plans. The *customer evaluation survey* assesses public confidence in the Judiciary every four years. The 'customers' are litigants, members of the Bar, the Public Prosecution Service and other repeat-players. The various groups of court users are asked about several aspects of the service provided by a court including for example how the judge interacts with litigants, the readability and comprehensibility of decisions and whether the hearings start on time. Courts also often make use of customer panels in order to examine the results of the surveys in more depth.

In addition, once every four years the courts conduct a *survey of staff satisfaction*. In such a survey, the staff are asked 'to evaluate the scope for personal development, the performance of those in charge, the court management board and the variety of work. The findings are used to adopt measures to improve staff motivation and satisfaction [… and] the performance of the organisation as a whole' (Council for the Judiciary 2008, p. 10). Again on a four yearly basis, a *committee visits* the courts. The committee members are 'outsiders' and can include for example a mix of university professors, lawyers, public prosecutors etc. The judicial performance measuring system includes also audits to assess progress on various aspects of the regulations. During an audit, the 'court staff examine whether the court meets the criteria included in the measuring system' (Council for the Judiciary 2008, 10).

Article 13 of the Order on the Financing of the Administration of Justice states that the outcomes of the quality system should determine the cost per minute set every three years by the Minister of Justice, even though it is not clear how this can occur. However, as noted by Philip Langbroek 'The way in which these considerations actually do lead to minute pricing is not entirely clear, as the outcomes are based on an immense flow of information from the courts to the Council to the Ministry, and on 'budget-negotiations' between the ministry and the Council for the judiciary […] However that may be, the outcomes of the quality measurement play some but not explicitly determined role in the budget and planning & control cycle of the courts and of the judicial organisation at large' (Langbroek 2010, p. 87).

4. From algorithms to politics

4.1. Algorithms

Mathematics, statistics and economics do matter, and they contribute to judicial performance evaluation. Their infiltration into the temple of the law is necessary and overdue. Quantitative measures are needed to evaluate specific areas of judicial performance and promote important values such as efficiency and timeliness. With the support of up-to-date case management systems, such measures can provide precise and effective analytical tools for measuring timeliness, efficiency and judicial effort. Contemporary judicial evaluation is, however, a complex and multifaceted exercise, involving more than just productivity, efficiency and timeliness. Values can be in competition, as can the frameworks or theories through which performance may be evaluated.

In some cases this competition results in a zero sum game where, for example, pursuing increased efficiency puts at risk key judicial values as fairness. As noted above, the introduction of Spanish *módulos* raised various problems. The studies of Domenech Pascual (2008) and of Bagues and Esteve-Volart (2010), and the

Oñati Socio-legal Series, v. 4, n. 5 (2014), 1099-1116
ISSN: 2079-5971

critiques of some of the Spanish judges associations[9], unveil a potential risk: that the modest increase in productivity generated by the system is the result of gaming and 'cherry picking'. As argued by a member of the Spanish judiciary (González Rivas, quoted by Domenech Pascual 2008, p. 15), the *módulos* 'push judges to prioritise the tasks and cases that are better rewarding in terms of 'points' and postpone those that are less rewarding, even if they constitute a fundamental guarantee of the rights of the citizens'. In this way, the *módulos* put at risk key judicial values: the judge is alone with the measurement and the goal. While the goal is just a means to an end (the overall justice performance), the incentive system pushes the judge to reach that goal and not the final end. Judges are given incentives to play with the system to reach the expected productivity, not to work for better overall judicial performance and justice service provision. This is one of the consequences of having unidimensional judicial performance evaluation systems measuring and promoting just one specific set of values, in this case managerial or economic. The second weakness of the *módulos* is the automatism between the measures and the incentive system. The existence of such an automatism, based on a simple mathematical algorithm, does not provide the occasions for joint reflection and collective evaluation, further increasing the risk of misplacement of means and ends. As a result the *módulos*, as with any other unidimensional and automatic judicial evaluation mechanisms, risk privileging one value over the others, and enact purely goal-oriented strategies.

There are many examples of judicial performance evaluation that are much more balanced (i.e. not unidimensional) and do not invoke automatic consequences based on the data gathered through the evaluation. These exercises often try to strike a balance between competing values (how much efficiency and how much fairness). An example is Courtools, which aims to offer a balanced view of court performance by establishing goals for effective court performance in five areas: access to justice, expedition and timeliness; equality, fairness, and integrity; independence and accountability; and public trust and confidence (Ostrom et al. 2005). No doubt CourTools, providing a balanced view of court performance, and requiring the involvement of human agents in the interpretation of the data collected, largely reduces the risks of the *módulos*. But improvements are even more desirable than a balanced view of court performance. While the concept of balancing implicitly entails that one performance area (or one value) is ameliorated at the detriment of other values (more fairness less efficiency), improvement means that amelioration is pursued without necessarily implying detriment in other areas (more efficiency and more fairness). Therefore, judicial performance evaluation should be addressed to support positive sum games, so that, through better organisational settings, procedural design, agreements with stakeholders and technology deployment different sets of values can be upheld in parallel: more efficiency and more fairness, more procedural justice and better effectiveness, and so on. The experience of various judiciaries, such as Sweden (Hagsgård 2008), Finland (Svela 2006), and the Netherlands, point to informed dialogue as a means for balancing and searching for improvements. So while the Dutch system, for example, quantifies the effort and cost of cases through the *Lamicie* model, this data is not used directly but only as the basis for the negotiation of the allocation of financial resources. This negotiation takes place through a two-tiered system, from the Ministry of Justice to the Council and from the Council to the courts, allowing for the infusion of non-economic values into the process, with the support of more qualitative and legal oriented assessments. The result is a system struggling to improve the quality of the justice service, of which productivity is just one component.

The Dutch case, being informed by a variety of methods and theoretical frameworks, allows identification of another issue associated with judicial

[9] See in particular http://forojudicial.es/wpfji/?tag=modulos

performance evaluation. As we have seen, the model is based on a dualism between *Lamicie* (quantitative and economic driven) and *RechtspraaQ,* based on a composite set of scientific frameworks (sociology, training, quality regulations etc.). The question is how *RechtspraaQ* can inform the budgeting process, and how *Lamicie* and the budgeting system can affect the judicial values promoted by *RechtspraaQ*. Langbroek has noted that 'the outcome of quality measurement [*RechtspraaQ*] have some but not explicitly determined role in the budget and annual planning & control cycle of the court and of judicial organisation at large' (Langbroek 2010, p. 141). This is not surprising because *Lamicie* and *RechtspraaQ* tackle two incommensurable aspects. It is useless or even dangerous to create an algorithm (or any formalised rule) establishing automatic and close-coupled connections between the two systems.

4.2. Incommensurability

As we have seen, the new forms of evaluation of judicial performance based on values such as efficiency and timeliness tend to focus on measurable areas: number of cases, number of judges, time etc. This in turn is linked to another easily quantifiable element: money, be it in terms of remuneration (increase or decrease) at individual level, or in terms of budget allocated to the court at organizational level. At the same time, the performance of the justice system is not limited to such easily quantifiable elements. Quoting a well-used aphorism attributed to Einstein, which is particularly relevant in judicial evaluation, 'not everything that can be counted counts, and not everything that counts can be counted'.

Nevertheless, these elements that cannot be easily counted such as quality of decision, independence etc. also need to be taken in to account when evaluating the overall performance of the judges, courts and judiciaries, and not just efficiency and timeliness. This is even truer when evaluations driven by the economic framework generate distorting effects, some of which can be visible only looking through different theoretical frameworks (i.e. legal or sociological). This is the case, for example, in the previously mentioned 'cherry picking' practice induced by the *módulos*. The distorting effects on productivity (such as the reduction of the number of judges producing more than 160%) are visible through the economic-quantitative evaluation lens. However, a 'legal lens' must be used to observe and analyse the impact of the prioritization of cases on the fundamental rights of the citizens. One question could be how to reduce the risk of distorting effects generated by the introduction of the economic models for the evaluation of the system? At the same time, distorting effects can also be generated by evaluations conducted from a legal perspective. Neglecting values like efficiency, these provide a fertile ground for the never ending judicial proceedings still affecting several judiciaries (Velicogna 2011, CEPEJ 2012). Again, part of these distorting effects cannot be assessed through the legal lens but must be considered as one of the risks of unidimensional evaluation systems.

We argue that multiple perspectives, such as legal, economic and social, must inform judicial evaluation, and that multiple evaluative frameworks, therefore, need to coexist. This leads us to a more general question, which is how to better manage the interaction of multiple evaluative perspectives and theoretical frameworks, being legal, economic, or sociological, in a comprehensive judicial evaluation endeavour. To deal with this issue, we can restate that algorithm and automatic mechanisms alone do not solve the problem. As Kuhn pointed out, the choice between theories 'cannot be resolved by proof' (Kuhn 2012, p. 74) or purely factual data. It can be reached only through other means such as persuasion, argument and counter-argument or, in some cases, a determination imposed by authority or fiat. This also affects judicial evaluation. Courts and judiciaries are characterised by components and features that can be described and analysed in quantitative terms, while others can be grasped only through qualitative analysis. The relationship between quantitative measures and qualitative assessments carried out through the

different theoretical frameworks or scientific disciplines therefore needs to be confronted.

The issue, here, is how to translate an economic based evaluation into something relevant and meaningful also in legal or social realms and, on the other hand, how to give relevant components of law, political and social paradigms weight in economic decisions. The difficulties increase as the 'proponents of different theories are … like native speakers of different languages' (Kuhn 2012, p. 85).

In the end, though, we need neither to compare concepts, objects and data that cannot be compared, nor to decide if a theoretical framework is more relevant or 'superior' to another. As in scientific discourse, judicial evaluation can take advantage of 'persuasion, argument and counter-argument' between the relevant players. Judicial evaluation needs to find a political synthesis, and to keep open the decision, the decisional process, and the process of evaluation. Humanness, discretional power, judgement and bias are unavoidable components of this synthesizing process. For this reason the accountability of the evaluation process is important. On the other hand, we cannot ask economics and law (as alternative paradigms) to do what they cannot do: generate an objective rank based on objective criteria, based on comparable measures coming from the various scientific domains. Objectivity cannot exist in a composite field such as the third branch, so wisdom is needed. We must combine human judgements with some discretionary power, and elements of subjectivity.

4.3. Politics

The political dimension is also relevant from another perspective. Judicial evaluation is not a merely technical exercise, where scientific method is applied to measure aspects of judicial or organisational behaviour, where algorithms are used to automatically rank judges, whose salaries are set according to those rankings, or to allocate resources to courts. The introduction of scientific methods cannot deny the complexity of the operation of justice, and the texture of values supporting its institutional mission. Judicial evaluation, in the final analysis, is a matter of which values have to be infused into judicial behaviour and court operation. Indeed, judicial evaluation relies upon the 'authoritative allocation of values' which, for Easton, is the essence of politics (Easton 1953, p. 129). Politics in this sense is largely a matter of dialogue, and a space for dialogue is needed for balancing values or enacting positive sum games.

Performance evaluation must recognise the importance of a range of values.[10] It must also take account of the authority that is entitled to define the values to be pursued. And, third, it must work through the consequences attached to evaluations: by what process, through which means and with what balance resource allocation decisions are made. If judicial evaluation also has a strong political connotation, in the terms we have identified, then effort has to be addressed to the means for a political discourse to take place.

The two cases discussed in this work make clear that dialogue is needed to uphold and balance different values. Individual performance evaluation associated with automatic mechanisms, as the Spanish case, denies any space for dialogue. The judge is left alone in a monologue about if and how to reach the goal and respond to the incentive system. However, even at individual level dialogue is possible if, for example, the data gathered through the evaluation schema are discussed and analysed with colleagues. Once more we can consider the Netherlands, where individual judges are evaluated on the bases of regular interviews conducted by the management board (de Lange 2012) composed of the court president, vice-

[10] We have elsewhere identified three main perspectives: legal, managerial and public or democratic (Contini and Mohr 2008, Mohr and Contini 2014). [2014 reference added to the list]

president and court manger. Such interviews are addressed to improvement through coaching or peer review (Langbroek 2005, p. 175). The periodical evaluation of individual judges is transformed into a learning opportunity, and judges' performance can be evaluated from a broad perspective (not just productivity). However, the feedback built into such a system remains mainly at the individual level, while consequences of any evaluation may also affect shared organisational features. Good evaluation mechanisms should help judges to improve as both individual decision makers and as organisational actors. As noticed in the introduction, the functioning of a court is more and more the result of a complex organisational system. Therefore, individual performance must include the evaluation of the judge not just as individual decision maker but also as organisational actor, i.e. for the broader set of contributions provided to the functioning of the court: deciding cases, promoting innovation, coaching or substituting for colleagues, just to mention a few. In this case, evaluation outcomes, including consequences, need to be negotiated in an organisation-wide process. Improvements can be achieved through a stronger individual effort by 'underproductive' judges, but also through a number of other measures taken at organisational level.

Scientific methods can be used to build objective evaluations, but turn out to require the allocation of values through political discourse if they are to become fruitful. It is only through political discourse that we may move from the zero sum game of algorithms developed under a single theoretical framework, with limited sets of values, to positive sum games built on dialogue between relevant players, theories and values, whose result needs not to be a choice of one value over another but a viable synthesis.

References

Bagues, M., and Esteve-Volart, B., 2010. *Performance Pay and Judicial Production: Evidence from Spain* [online]. Unpublished manuscript. Available from: http://www.manuelbagues.com/bagues%20%26%20esteve-volart%20-%20judges.pdf [Accessed 20 December 2014].

CEPEJ, 2012. *European judicial systems, edition 2012 (2010 data)* [online]. Strasbourg: Council of Europe. Available from: http://www.coe.int/t/dghl/cooperation/cepej/evaluation/2012/Rapport_en.pdf [Accessed 13 December 2014].

Consejo General del Poder Judicial, 1999. *Measuring the Workload of Courts and Judges: The Spanish Approach* [online]. Madrid: Consejo General del Poder Judicial, excerpt from Memoria, Volumen I. pp. 257-275. English translation of pp. 257-275. Available from: http://siteresources.worldbank.org/INTLAWJUSTINST/Resources/SpanJudgesPerfIndicators.pdf [Accessed 21 December 2014].

Contini, F., and Mohr, R., 2007. Reconciling independence and accountability in judicial systems. *Utrecht Law Review* [online], 3 (2), 26-43. Available from: http://www.utrechtlawreview.org/index.php/ulr/article/view/46/46 [Accessed 21 December 2014].

Contini, F., and Mohr, R., 2008. *Judicial Evaluation. Traditions, innovations and proposals for measuring the quality of court performance*. Saarbrücken: VDM.

Council for the Judiciary, 2004. *The judiciary system in the Netherlands: a single autonomous organization*. The Hague: The Council for the Judiciary.

Council for the Judiciary, 2008. *Quality of the judicial system in the Netherlands* [online]. The Hague: Netherlands Council for the Judiciary. Available from: http://www.rechtspraak.nl/English/Publications/Documents/Quality-of-the-judicial-system-in-the-Netherlands.pdf [Accessed 21 December 2014].

Council for the Judiciary, 2013. *The Financing System of the Netherlands Judiciary* [online]. The Hague: Council for the Judiciary. Available from: http://www.rechtspraak.nl/English/Publications/Documents/The-Financing-System-of-the-Netherlands-Judiciary.pdf [Accessed 21 December 2014]..

de Lange, Roel. 2012. Judicial Independence in The Netherlands. *In*: A. Seibert-Fohr, ed. *Judicial Independence in Transition*. Heidelberg: Springer, 231-271.

Decker, K., Möhlen, C. and Varela, D.F., 2011. *Improving the performance of justice institutions. Recent experiences from selected OECD countries relevant for Latin America* [online]. Washington, D.C.: The World Bank. Available from: http://siteresources.worldbank.org/INTECA/Resources/librojusticiaING-cian.pdf [Accessed 21 December 2014].

Domenech Pascual, G., 2008. La Perniciosa Influencia de las Retribuciones Variables de los Jueces sobre el Sentido de sus Decisiones. *Indret: Revista para el Anlisis del Derecho* [online], 3, Article 8. Available from: http://www.indret.com/pdf/569_es.pdf [Accessed 21 December 2014].

Easton, D., 1953. *The Political System: An Inquiry into the State of Political Science*. New York: Knopf.

Hagsgård, M.B., 2008. Internal and external dialogue: A method for quality court management. *International Journal For Court Administration* [online], 1 (2), 10-18. Available from: http://www.iacajournal.org/index.php/ijca/article/view/98/79 [Accessed 13 December 2014].

Kuhn, T.S., 2012. Objectivity, Value Judgement, And Theory of Choice. *In*: A. Bird and J. Ladyman, eds. *Arguing about Science*. New York: Routledge, 74-86.

Langbroek, P., 2005. Recruitment, Professional Evaluation and Career of Judges and Prosecutors in the Netherlands. *In*: G. Di Federico, ed. *Recruitment, Professional Evaluation and Career of Judges and Prosecutors in Europe*. Bologna: Lo Scarabeo, 156-86.

Langbroek, P., ed., 2010. *Quality Management in Courts and in the Judicial Organisations in 8 Council of Europe Memberstates*. Utrecht: Montaigne Centre and CEPEJ.

Mohr, R., Contini, F., 2014. Conflicts and Commonalities in Judicial Evaluation. *Oñati Socio-legal Series* [online], 4 (5), 843-862. Available from: http://ssrn.com/abstract=2537860 [Accessed 23 December 2014].

Ng, G.Y., 2007 *Quality of Judicial Organisation and checks and balances* Antwerp: Intersentia.

Ostrom, B.J., et al., 2005. *CourTools: Trial Court Performance Measures*. Williamsburg, Va.: National Center for State Courts.

Poblet, M., and Casanovas, P., 2005. Recruitment, Professional Evaluation and Career of Judges and Prosecutors in Spain. *In*: G. Di Federico, ed. *Recruitment, professional evaluation and career of judges and prosecutors in Europe*. Bologna: Lo Scarabeo, 185-214.

Signifredi, P., 2006. *Misurare la produttività dei giudici: il caso Spagnolo*. Bologna: IRSIG-CNR.

Svela, A., ed., 2006. *Evaluation of the quality of the adjudication in courts of law: principles and proposed quality benchmarks*. Oulu: Painotalo Suomenmaa.

Tribunal Supremo, 2012. *Sentencia T.S. (Sala 3) de 8 de febrero de 2012* [online]. Available from: http://portaljuridico.lexnova.es/jurisprudencia/JURIDICO/120996/sentencia-

ts-sala-3-de-8-de-febrero-de-2012-cgpj-nulidad-de-acuerdo-de-la-comision-permanente [Accessed 21 December 2014].

UNODC, 2011. *Resource Guide on Strengthening Judicial Integrity and Capacity* [online]. Vienna: United Nations. Available from: http://www.un.org/zh/issues/anti-corruption/pdfs/resource_guide.pdf [Accessed 21 December 2014].

van der Torre, A., et al., 2007. *Rechtspraak: Productiviteit in perspectief* [online]. Den Haag: Rechtspraak Sociaal en Cultureel Planbureau. Available from: http://www.rechtspraak.nl/Organisatie/Raad-Voor-De-Rechtspraak/WetenschapsOnderzoek/publicatieswetenschappelijkonderzoek/Documents/Rechtspraak-productiviteit-in-perspectief.pdf [Accessed 21 December 2014].

Velicogna, M., 2011. Study on Council of Europe Member States on Appeal and Supreme Courts' Lengths of Proceedings. *CEPEJ Studies* [online], 17. Available from: https://www.academia.edu/9860581/Study_on_Council_of_Europe_Member_States_on_Appeal_and_Supreme_Courts_Lengths_of_Proceedings_2011_ [Accessed 22 December 2014].