



Innovating Judicial Performance Evaluations: Toward Academic-Style Peer Review?

SILVIO ROBERTO VINCETI 

ACADEMIC ARTICLE



ABSTRACT

Evaluating judicial performances is a crucial, albeit often underappreciated element of judicial governance, as it incentivizes judges' productivity and allows for meritocratic career advancement within the judiciary. At the same time, judicial performance evaluation remains a daunting task, plagued by different recurring issues. Alongside the controversial nature of judicial merit and conflicting expectations as to their ultimate goal, methods of judicial evaluation are often saddled with specific technological limitations, such as the risk of cognitive bias and data intelligibility. To overcome these shortcomings, it may be useful to start thinking of judicial performance in terms of other enterprises where merit is similarly hard to assess and measure objectively. A relevant example is the comparison with scholarly work, which resembles judicial activity in that both are goal-oriented enterprises where quality assessment proves somewhat elusive. Cognizant of the difficulties, however, researchers and academic institutions have long developed tools to overcome limitations. Associative thinking with scholarly research could thus yield positive ideas for innovating the assessment of judges' work, one prime example being the import of scholarly blind "peer" review processes to judicial performance evaluations.

CORRESPONDING AUTHOR:

Dr. Silvio Roberto Vinceti

Research Fellow, University
of Modena and Reggio, Italy

silvioroberto.vinceti@unimore.it

KEYWORDS:

judicial performance;
judicial performance
evaluation; peer review;
scientific methods;
comparative judicial studies

TO CITE THIS ARTICLE:

Silvio Roberto Vinceti,
'Innovating Judicial
Performance Evaluations:
Toward Academic-Style
Peer Review?' (2024) 15(1)
*International Journal for Court
Administration* 2. DOI: [https://
doi.org/10.36745/ijca.576](https://doi.org/10.36745/ijca.576)

At a very basic level, judicial performance evaluation falls into place as a chapter of job performance appraisal. Just as companies are concerned with fostering their employees' abilities and enhancing their competence, states have a vested interest in having their judges fulfill their duty with increasing rigor and proficiency. To the extent that the "cognitive process is universal,"¹ what holds of any job should roughly apply to the judicial office as well: Work evaluations are thus expected to increase judicial performances both in the sense that judges take stock of the feedback they receive and because of the more general and somewhat unconscious phenomenon by which "the mere fact of knowing that one is being observed or measured increases performance and fosters cooperative behavior."² In this respect, the assessment of judicial performances very much resembles the evaluation of any public servant. By contrast, members of parliaments and political actors in general are not subjected to analogous kinds of performance evaluations since popular elections are understood to be the only legitimate method for the evaluation of political agency.³

Besides the natural interest in the improvement of the judges' competence, there seems to be something peculiarly distinctive to judicial performance evaluations. From Ancient Rome⁴ to the present day,⁵ a pyramidal grade structure marks the history of the judicial profession. Although the strictest hierarchical aspects may have been relinquished in time also in light of concerns for internal judicial independence, the judiciary remains a graded organization in which higher roles come in scarce quantity. For this reason, it is all the more natural that different individuals may apply and compete for the very same positions, thus begging the question of how to select the best profiles to fill the vacancies. Absent objective tools for evaluating judges, informal rules would reign in career advancement, with increased risk of cronyism,

1 Dina Van Dijk and Michal M Schodl, 'Performance Appraisal and Evaluation' in James D Wright (ed), *International Encyclopedia of the Social & Behavioral Sciences (Second Edition)* (Elsevier 2015) 717 <<https://www.sciencedirect.com/science/article/pii/B9780080970868220345>> accessed 12 May 2023.

2 *ibid* 716.

3 It is no coincidence that political scientists name "performance voting" the idea by which "voters, when they cast their vote, evaluate government performance and vote according to this evaluation: Voters holding a positive evaluation will be more likely to vote for an incumbent party, whereas voters evaluating incumbent performance negatively are more likely to support the opposition." Dieter Stiers, 'Performance Voting, Retrospective Voting, and Economic Voting. Conceptual Clarity and Empirical Testing' (2022) 103 *Social Science Quarterly* 399, 399.

4 Comparisons between current judiciaries and Roman magistrates are notoriously difficult because of the blurred lines between law-creation and law-application in the *iurisdictio* function at that time. See Alberto Burdese, 'Voce Magistrato (*Dir. Rom.*)' 187. However, it is symptomatic that judicial functions were part of the *cursus honorum*, which to this day epitomizes the idea of career progression in public service: cf Eric A Posner, 'The Constitution of the Roman Republic' in Giuseppe Dari-Mattiacci and Dennis P Kehoe (eds), *Roman Law and Economics: Institutions and Organizations Volume I* (Oxford University Press 2020) 52 <<https://doi.org/10.1093/oso/9780198787204.003.0003>> accessed 14 May 2023.

5 Cf Giuseppe Di Federico (ed), *Recruitment, Professional Evaluation and Career of Judges and Prosecutors in Europe: Austria, France, Germany, Italy, the Netherlands and Spain* (Lo scarabeo 2005).

favoritism, chauvinism, and improper political influence.⁶ In this respect, the very possibility that the judiciary can be framed as a meritocratic institution is premised on the assumption that we can somewhat objectively tell a superior judge from a mediocre or even an incompetent one.

Concerns over distorted career advancement in the judiciary are not abstract hypotheticals, as recent events in the Italian experience recently demonstrated. In Italy, as it happens, judicial evaluations have long been considered problematic insofar as all judges (>99%) receive the very same judgment — “positive” — in their periodical professional evaluations.⁷ To the extent prospective candidates fare all equally well on paper, there have been suspects that judicial appointments and promotions may at times be explained more by the candidate’s affiliation to a particular judicial association, and the association’s sway over the current composition of the Superior Council of the Magistracy (SCM), than by the judge’s objective abilities.⁸ The eruption of the so-called “Palamara affair” in 2019 uncovered a system in which promotions and appointments appeared the objects of a power struggle between different associations, thus proving that fears of cronyism in judicial career advancement were not utterly misplaced.⁹ In some cases linked to the affair, it seemed as though the judge’s personal relations to the reigning *siloviki* in the SCM could be factored as the main determinant for her career advancement. Unsurprisingly, Giuseppe Di Federico, a renowned scholar and expert on the Italian judiciary, pointed at the current state of judicial performance evaluations as one fundamental reason for the scandal,¹⁰ and in the wake of the affair the Italian Parliament eventually attempted to address the

6 “Firstly, performance evaluation can be a powerful tool to aid the accountability mechanism that is judicial promotion. This role is particularly important for the career judiciaries of the continental judicial systems, where there is a clear imperative for a system to select judges to populate the higher ranks of the judiciary. Appointment on the basis of purely arbitrary or nepotistic grounds is unacceptable. Given the normative impact of higher courts decisions, it is clearly preferable to populate the higher judiciary with the best judges.” Joe McIntyre, ‘Evaluating Judicial Performance Evaluation: A Conceptual Analysis’ (2014) 4 *Oñati Socio-Legal Series* 898, 910.

7 Giuseppe Di Federico, ‘Magistrati ordinari: Organico, Reclutamento, Valutazioni della professionalità, Carriera, Tramutamenti e attribuzione delle funzioni giudiziarie, Trattamento economico’ in Giuseppe Di Federico (ed), *Ordinamento giudiziario: uffici giudiziari, CSM e governo della magistratura* (Bononia University Press 2019) 343–347; Simone Benvenuti, ‘The Italian System of Judicial Governance: An Arena of Confronting Informal Practices and the Push Towards Formalization’ (2023) 24 *German Law Journal* 1373, 1387–1388; Marco Fabri, ‘Clash of Visions: Regulating Judges and Prosecutors in Italy’, *Regulating Judges* (Edward Elgar Publishing 2016) <<https://www.elgaronline.com/edcollchap/edcoll/9781786430786/9781786430786.00018.xml>> accessed 22 April 2024. The same happens in Romania: see Lukáš Hamřík, ‘Actors of Informal Judicial Institutions and Practices’ (2023) 24 *German Law Journal* 1520, 1530.

8 See Giuseppe Di Federico, ‘Riforma delle valutazioni di professionalità e delle disfunzioni ad essa connesse, in prospettiva comparata’ [2022] *Diritto di difesa* 1, 21–23.

9 On the “Palamara affair” see Gabriella Mangione, ‘Some Brief Remarks on the Controversial Relationship Between the Judiciary and Politics in Italy’ (2021) 27 *Comparative Law Review* 79, 80; Maurizio Catino, Cristina Dallara and Sara Rocchi, ‘The Organizational Reasons for Wrongdoing. The Case of Italy’s Superior Council of the Judiciary (CSM)’ (2023) 79 *Crime, Law and Social Change* 453, 458; Silvio Roberto Vinceti, ‘Standardisation and Authority in Judicial Discipline: A Comment on Italy and Ireland’s Experiences’ (2022) 6 *Irish Judicial Studies Journal* 55, 56; Alexandru Olanescu, ‘Discretionary Power and Abuse of Power in the Activity of State Institutions’ (2021) 84 *Curentul Juridic* 80, 90–91.

10 Angela Stella, ‘Intervista al professor Di Federico: “Grazie a Palamara hanno fatto carriera anche i giudici che lo giudicheranno”’ *Il Riformista* (22 July 2020) <<https://www.ilriformista.it/intervista-al-professor-di-federico-grazie-a-palamara-hanno-fatto-carriera-anche-i-giudici-che-lo-giudicheranno-134353/>> accessed 14 May 2023.

situation by reforming judicial performance evaluations.¹¹ Other judiciaries seem to display quite comparable dynamics.¹²

By contrast, in other countries, the absence of a fair and objective evaluation of judges lent itself to a different kind of problem. In Eastern Europe, for example, scholars have highlighted the use of judicial evaluations as a tool for undue influence on the administration of justice.¹³ The connection between judicial evaluation and the independence of the judiciary cannot be overstated. To the extent that judges need positive appraisals either to be promoted or to avoid reprehension — be it disciplinary or not — judicial evaluations represent an obvious means to exert influence on the judiciary and the application of the law more generally. The independence of judicial evaluators thus clearly affects judicial independence per se. In this respect, the possibility of objectivity in judicial evaluation is all the more important as it allows for external review to keep in check both inward favoritism and outward political influence in the assessment and promotion of judges.

And yet, despite the salience of these issues and the ever-aggrandizing role of judiciaries in contemporary societies,¹⁴ judicial performance evaluations have struggled to elicit the same kind of attention reserved, for example, to judicial selection or judicial discipline. While there are notable exceptions,¹⁵ the persistence of practical issues¹⁶ and conceptual ambiguity as to their ultimate goal¹⁷ suggest that judicial performance evaluations remain a somewhat underresearched topic, especially when compared with other traditional issues in judicial governance. The aim of this short paper is thus both to draw renewed attention to the topic and to propose one conceptual avenue that could hopefully address some ongoing limitations in current methods of judicial performance evaluation. The main idea is to show that judicial performance evaluations arguably benefit from associative thinking with other human enterprises for which quality assessment has long appeared a complicated task and, namely, academic scientific scholarship. An example of one possible development

11 In delegating the government, the power to adopt legislative decrees to the reform of the judicial system the law no. 71/2022 explicitly contemplates the reforms of judicial evaluation, but quite surprisingly limits the possibility of grading judicial performance to «the judge's ability to organize its own work». See Legge 17 giugno 2022, n. 71 — Deleghe al Governo per la riforma dell'ordinamento giudiziario e per l'adeguamento dell'ordinamento giudiziario militare, nonché disposizioni in materia ordinamentale, organizzativa e disciplinare, di eleggibilità e ricollocamento in ruolo dei magistrati e di costituzione e funzionamento del Consiglio superiore della magistratura art. 3.1.c.

12 Nino Tsereteli, 'Constructing the Pyramid of Influence: Informal Institutions as Building Blocks of Judicial Oligarchy in Georgia' (2023) 24 German Law Journal 1469, 1484–1485.

13 Denis Preshova, Ivan Damjanovski and Zoran Nechev, 'The Effectiveness of the "European Model" of Judicial Independence in the Western Balkans: Judicial Councils as a Solution or a New Cause of Concern for Judicial Reforms' [2017] Cleer Papers 7, 25; Attila Vincze, 'Schrödinger's Judiciary—Formality at the Service of Informality in Hungary' (2023) 24 German Law Journal 1432, 1440.

14 Cf C Neal Tate and Torbjörn Vallinder (eds), *The Global Expansion of Judicial Power* (New York University Press 1995).

15 See eg Francesco Contini (ed), *Handle with Care Assessing and Designing Methods for the Evaluation and Development of the Quality of Justice* (IRSIG-CNR 2017) <https://www.academia.edu/36129389/Handle_with_Care_Assessing_and_designing_methods_for_the_evaluation_and_development_of_the_quality_of_justice> accessed 14 May 2023.

16 See *infra* Part II.A.

17 See *infra* Part II.B.

arising out of the conceptual association is the implementation of double-blind peer review mechanisms in judicial performance evaluations.

In Part I I briefly describe the two main approaches for measuring judicial performance and assessing judicial merit: the use of surveys and the operationalization of objective data regarding adjudication. While they do not exhaust the catalog of possible methods, the two approaches seem the most relevant within the field of the “objective” approaches of judicial evaluation — as opposed, for example, to judicial elections. In Part II two main criticisms concerning judicial performance evaluations are described. On the one hand, the judicial process seems to differ from the standardizable type of action that is normally operationalized in performance appraisal. On the other, judicial performance evaluation seems at times simultaneously charged with the tasks of both assessing professional standards among judges and evaluating judicial excellence. While both goals can be pursued within a given system, it is doubtful that the same tool can serve both aims. In Part III I eventually discuss the association of judicial performance evaluation to academic scholarship — the “Academic Parallel” — and its possible applications.

I. AN OVERVIEW OF SOME METHODS OF JUDICIAL PERFORMANCE EVALUATION

A. SURVEYS & INTERVIEWS

The use of surveys goes to the origins of judicial performance evaluation. The first example is the well-known Chicago Bar Association’s surveying of judges’ performances in 1873 to inform voters before judicial elections.¹⁸ With the spread of similar surveys throughout the United States the American Bar Association eventually took a role in standardizing and “professionalizing” the assessment method¹⁹ and some researchers thus consider Alaska’s 1976 state-sponsored surveying program as the beginning of *modern* judicial performance evaluation.²⁰ The class of typical respondents to these surveys now generally comprises a wide range of individuals who may be cognizant of the judge’s work, from colleagues to court staff.²¹ Besides other common law systems,²² the practice of surveying stakeholders for judicial performance evaluation also spread through civil law countries: For example, the Netherlands’ judicial quality system, called “RechtspraakQ”, utilizes both client evaluation surveys and staff satisfaction surveys for judicial evaluations.²³

To the extent that they partake in a dynamic where at least the evaluator knows the identity of the evaluatee, interviews can be conflated with surveys. A prominent case of interview-based judicial evaluations can be found in the French judiciary.

18 Jennifer Elek and David Rottman, ‘Methodologies for Measuring Judicial Performance: The Problem of Bias’ (2014) 4 *Oñati Socio-Legal Series* 863, 867.

19 *ibid* 867–868.

20 McIntyre (n 6) 902.

21 See eg Rebecca White Berch and Erin Norris Bass, ‘Judicial Performance Review in Arizona: A Critical Assessment’ (2014) 4 *Oñati Socio-Legal Series* 927, 933.

22 See eg Marilyn Warren, ‘From Evaluation to Improvement: A Chief Justice’s Perspective’ (2014) 4 *Oñati Socio-Legal Series* 953, 964–968.

23 See eg the “RechtspraakQ” system in the Netherlands Francesco Contini and others, ‘Formula over Function? From Algorithms to Values in Judicial Evaluation’ (2014) 4 *Oñati Socio-Legal Series* 5, 1099, 1109–10.

Every year — biannually in the case of administrative courts — the head of the court interviews the judges in her Court.²⁴ The discussion is conducted in light of written notes submitted by the judges themselves, in which they recount their professional activities and the training processes they attended. The head of court eventually writes up a summary detailing the content of the interview, which also takes into consideration observations from other fellow judges on the evaluatee's behavior and abilities. Besides the paradigmatic France experience, interview-based methods of judicial performance assessment can also be found in several States of the United States²⁵ and in Sweden.²⁶

B. OBJECTIVE DATA COLLECTION

While surveys and interviews can be considered data for the purposes of judicial performance evaluations, they remain a human-made product that intuitively contrasts with the kind of objectified data that can be extracted from other measurable aspects of judicial activity. A second fundamental avenue for the performance evaluation of judges lies in the collection of quantitative data concerning adjudication. As plenty of aspects within the judicial process already come broken down in pure numbers — from the time needed to dispose of cases to the ratio of judgments reversed by higher courts — it may strike as a convenient route to make use of these readily available quantitative data to determine the performance of the particular judge supervising a case through mathematical algorithms.²⁷

Two of the most frequently employed types of data are disposition time and reversal rates. Disposition time is the “commonly used indicator to estimate the time a judicial system takes to resolve a case,”²⁸ while reversal rate can be defined as “[t]he number of times the judge has been reversed by a higher court during the evaluation period divided by the number of cases appealed.”²⁹ A prominent example of the use of disposition time can be found in Spain, where since 2000 a part of the judges' remuneration depends on their productivity, which is deduced from the comparison of the time the judge needed to complete judicial tasks and the expected time, as standardized in specific measurement tables (*módulos*).³⁰ A similar system exists in

24 See Hélène Pauliat and others, 'The Evaluation and Development of the Quality of Justice in France' in Francesco Contini (ed), *Handle with Care Assessing and designing methods for the evaluation and development of the quality of justice (IRSIG-CNR 2017)* 89–90 <https://www.academia.edu/36129389/Handle_with_Care_Assessing_and_designing_methods_for_the_evaluation_and_development_of_the_quality_of_justice> accessed 14 May 2023.

25 Richard C Kearney, 'Judicial Performance Evaluation in the States' (1999) 22 *Public Administration Quarterly* 468, 474; Jordan M Singer, 'Judicial Performance Evaluation in the States: The IAALS JPE 2.0 Pre-Convening White Paper' [2022] *SSRN Electronic Journal* 13 <<https://www.ssrn.com/abstract=4505778>> accessed 22 October 2023.

26 Marie B Hagsgård, 'Internal and External Dialogue: A Swedish Approach to Quality Work in Courts' (2014) 4 *Oñati Socio-Legal Series* 993, 999–1001.

27 See eg Francesco Contini, Richard Mohr and Marco Velicogna, 'Formula over Function? From Algorithms to Values in Judicial Evaluation' (2014) 4 *Oñati Socio-Legal Series* 1099.

28 OECD, *Government at a Glance 2021 (Organisation for Economic Co-operation and Development 2021)* 236 <https://www.oecd-ilibrary.org/governance/government-at-a-glance-2021_1c258f55-en> accessed 22 October 2023.

29 Rebecca D Gill, 'Implicit Bias in Judicial Performance Evaluations: We Must Do Better Than This' (2014) 35 *Justice System Journal* 301, 319.

30 Contini, Mohr and Velicogna (n 27) 1103–1106.

the Netherlands.³¹ Reversal rates are also natural candidates for evaluating judicial performances.³² A general common-sense assumption is that the higher court's decision to overturn a lower court's sentence depends on some erroneous appreciation either of the law or of the facts of the case on the part of the lower judge, whereas a rejected appeal is understood as validation of the first instance judge's sound reasoning. The more cases are reversed, the less the judge is deemed proficient, and vice versa. Unsurprisingly, the use of reversal rates in judicial performance evaluations is common both in the United States³³ and Europe.³⁴

II. OPEN ISSUES IN JUDICIAL PERFORMANCE EVALUATION

A. OVERARCHING ISSUES AND TECHNOLOGICAL LIMITATIONS

The first and foremost difficulty in judicial performance evaluation boils down to the fact that judicial activity is hardly the kind of standardizable behavior to be smoothly operationalized in performance evaluation. A sprinter running a 100 m race, a call center agent consulting clients over the phone, a physician providing the patients with a diagnosis: Although all these human enterprises may entail some level of measurement subtleties, they are all amenable to familiar methods of quality assessment, be it the use of a chronometer, questionnaires on customers satisfaction, or complex tools employing diagnosis-related groups. By contrast, the process of solving disputes through the application of rules is recalcitrant to any straightforward method of objective evaluation. No one is more accustomed to the problem than judges' biographers who are forced to struggle with the question of what it is that made the subject of their book a "great judge" after all.³⁵

Concurring elements account for this difficulty in the rating of judges. First, it is not plain and obvious that "perfect lawfulness" should work as the "right normative standard" for judicial performance evaluation.³⁶ In other words, it is not clear whether the product of adjudication should be weighed against the best possible account of existing law, or rather evaluated in light of the beneficial effects it delivers for society as a whole or the satisfaction it brings to the parties in the case. As Lawrence Solum once

³¹ *ibid* 1106–1110.

³² Rebecca D Gill, Sylvia R Lazos and Mallory M Waters, 'Are Judicial Performance Evaluations Fair to Women and Minorities? A Cautionary Tale from Clark County, Nevada' (2011) 45 *Law & Society Review* 731, 743–744.

³³ Lloyd Musolf, 'Performance Evaluation of Federal Administrative Law Judges: Challenge for Public Administration?' (1998) 28 *The American Review of Public Administration* 390, 395; Malia Reddick and Rebecca Love Kourlis, 'Judicial Performance Evaluation' in Peter Koelling and Rebecca Love Kourlis (eds), *The improvement of the administration of justice* (American Bar Association 2016) 161.

³⁴ Anne Sanders, Report on the Individual Evaluation of Judges in Albania, 2014, 5, <https://rm.coe.int/eu-coe-support-to-efficiency-of-justice-sej-a-joint-project-between-th/168078874d>.

³⁵ Cf Willard Hurst, 'Who Is the "Great" Appellate Judge?' (1949) 24 *Indiana Law Journal* 394; Linda Przybyszewski, 'The Dilemma of Judicial Biography or Who Cares Who Is the Great Appellate Judge? Gerald Gunther on Learned Hand' (1996) 21 *Law & Social Inquiry* 135.

³⁶ See Gregory Mitchell, 'Evaluating Judges' in David E Klein and Gregory Mitchell (eds), *The Psychology of Judicial Decision Making* (Oxford University Press 2010) 231–234 <<https://doi.org/10.1093/acprof:oso/9780195367584.003.0014>> accessed 15 May 2023.

noted, in the Western, Aristotelian tradition of justice the “lawfulness conception” has often given way to a conception in which “equity” and “practical wisdom” could occasionally trump the letter of the law “if the rule led to consequences that were absurd or manifestly unjust.”³⁷ Judges themselves seem keen to treasure the importance of non-legal skills and abilities in measuring judicial excellence, as a study by Jennifer Elek made clear:

[W]hat “judicial excellence” means to judges appears to be much broader than what might be traditionally assumed. For example, judges in this study described the importance not only of knowledge of the law and court rules, policies, and procedures, but also of a wider body of practical and operational knowledge. They described this broader universe of knowledge about the court community, stakeholder agencies, and other resources as essential for effective problem-solving and decision-making. Notably, judicial interviewees highlighted the importance of so-called “soft skills”, such as interpersonal and emotion management skills, when discussing judicial excellence and the ability to perform judicial work most effectively.³⁸

But even if we were to subscribe to the “perfect lawfulness” account, issues of evaluating judicial performances would not wane overnight. Consider the widely known and epistemologically refined theory of originalism in the United States: Even sincerely committed originalists adopting the same methodological framework do come to different conclusions as to the meaning of some clauses of the US Constitution.³⁹ The disagreement is at least partly due to the acknowledged existence of open-ended clauses that are amenable to different interpretations and in turn to generating multiple legal rules.⁴⁰ While it may be possible to think that non-constitutional adjudication entails a lesser degree of controversy, at least some level of indeterminacy seems essential to the law in itself⁴¹ and to this extent, objective evaluation of judicial performances may occasionally result in an impossibility.

Besides the overarching problem of the idiosyncratic nature of adjudication as a human activity, some issues specifically affect each method of judicial evaluation. On the one hand, to the extent that the evaluator knows the evaluatee, surveys and interview-based methods are evidently prone to bias. For example, concerns of prejudice against women and minority judges have been voiced in the American

37 Lawrence B Solum, ‘A Tournament of Virtue Symposium: Empirical Measures of Judicial Performance’ (2005) 32 Florida State University Law Review 1365, 1383–1385.

38 Jennifer K Elek, ‘Judicial Perspectives on Emotion, Emotion Management, and Judicial Excellence in the USA’ (2019) 9 Oñati Socio-Legal Series 865, 872–873.

39 Think, eg., of the longstanding debate over whether *Brown v. Board of Education* was correctly decided on originalist grounds. Cf. Robert H Bork, *Tempting of America. The Political Seduction of the Law* (Simon & Schuster 1990) 74–84; Raoul Berger, *Government by Judiciary: The Transformation of the Fourteenth Amendment* (1977) (2nd edn, Liberty Fund 1997) 457.

40 Lawrence B Solum, ‘Originalist Methodology’ (2017) 84 University of Chicago Law Review 269, 294–295. A different but equally important explanation for disagreement among originalists is some persisting confusion between originalism as a “positive theory of interpretation” and “normative theory of action”: see Gary Lawson, ‘Equivocal Originalism’ (2022) 27 Texas Review of Law and Politics 309.

41 See Brian Leiter, ‘Legal Indeterminacy’ (1995) 1 Legal Theory 481.

experience.⁴² On the other hand, data collection can be affected by the unique nature of the legal process. The comparison of case disposal time, for example, may fail to grasp the specificities of particular proceedings and it is a common experience that some cases may be more time-consuming or intellectually challenging than others because of the number of parties involved, lengthy evidence-gathering technology, and so on.⁴³ The use of case reversal ratios is also problematic. First, it is sometimes possible that the reversal of the lower court decision may be due to the uncovering of new decisive evidence in appeal and not because of a defect in the original pronouncement: In such an instance, the lower judge's decision is neither right nor wrong for the purposes of the performance evaluation. Secondly, the problem with the use of case reversal rates lies in the scope of the methodology: Besides failing to acknowledge the different quality between two equally upheld decisions, reversal ratios are also useless in the case of judges sitting in courts of last resort.

B. PROFESSIONALISM VS EXCELLENCE

A second issue affecting judicial performance evaluations is a certain dose of ambiguity as to their ultimate goal. Once again, the Italian experience offers a compelling example concerning an argument voiced⁴⁴ to legitimize the current system of judicial evaluation.⁴⁵ As their very name of “professionalism evaluations” (*valutazioni di professionalità*) somewhat attests to, performance evaluations of Italian judges are not so much aimed at evaluating excellence as to make sure that all members of the judiciary meet a certain minimum threshold. While it is perfectly rational of a judicial system to both demand that all judges meet minimum levels of competence and to single out the best ones for promotion, it is at least counterintuitive to expect that the very same tool could *necessarily* serve both ends. In this regard, it is quite doubtful that judicial performance evaluations designed to assess minimum standards can also detect judicial excellence. Instead, what would be needed to single out the superior judges for promotion and directive roles is more akin to a ranking system along the lines of what Stephen Choi and Mitu Gulati advanced with the idea of a “tournament of judges.”⁴⁶

III. THE ACADEMIC PARALLEL

A. UNITED BY INCOMMENSURABILITY

At the outset, the task of the judge and that of the researcher may seem remarkably different. While the former is expected to discover how the sources of law bear on one case at issue, the latter is concerned with validating or refuting a particular

⁴² See eg Angela Melville, ‘Evaluating Judicial Performance and Addressing Gender Bias’ (2014) 4 *Oñati Socio-Legal Series* 880.

⁴³ For similar criticisms see Contini, Mohr and Velicogna (n 27).

⁴⁴ See *supra* Introduction.

⁴⁵ Unsurprisingly, this is what is pointed out by leaders of judicial associations: see Valentina Stella, ‘Intervista a Eugenio Albamonte. «Valutare Noi Toghe? No a Un’élite Di Eccellenti, Ma i Capi Vanno Formativi»’ *Il Dubbio* (24 March 2021) 1.

⁴⁶ Stephen Choi and Mitu Gulati, ‘A Tournament of Judges?’ (2004) 92 *California Law Review* 299.

thesis about some “state of affairs” in the world,⁴⁷ either in the familiar sense of a scientist that seeks to unveil the biological causes of a disease or in the more nuanced sense of a historian concerned with explaining the reasons of some specific event.⁴⁸ However, besides their eminently intellectual nature, what unites the two enterprises is that they remain strictly goal-oriented activities. If, for example, in the realm of the arts, the very purpose may well remain entirely subjective, both the judge and the researcher have “preassigned” institutional functions: While the former is concerned with uncovering the sources of law and the way they relate to the facts of the case at hand, the latter aims at discovering the truth about some particular state of affairs in the world. Whatever the different methodologies they may practically employ, it is at least theoretically clear what it is that the lawyer and the researcher are seeking.

For the purposes of this inquiry, however, the most relevant *tertium comparationis*⁴⁹ is the elusive nature of quality assessment in both the judicial and the scientific research domains. How can we tell a good researcher from an inferior one, a valuable scientific paper from a neglectable one? These questions have been raised time and again within academic institutions and the scientific community as a whole.⁵⁰ Our concern here is not with fundamental philosophical questions such as the nature of “scientific progress.”⁵¹ What we seek by associatively thinking about the two domains is to trade on the pre-existing and more developed debate over scientific research quality in order to see if any of the practical tools devised by academic institutions can be transplanted into the context of judicial performance evaluation. As we turn our attention to scientific research, we are soon made aware that “traditional peer review remains the gold standard for evaluating and selecting quality scientific publications.”⁵² We should thus ask ourselves if such a “gold standard” may have any bearing on judicial performance evaluations.

B. TOWARD DOUBLE-BLIND JUDICIAL PERFORMANCE EVALUATIONS?

It is beyond the scope of this paper to outline a comprehensive system for objective judicial evaluation based on the Academic Parallel. What is sought here, instead, is to discuss a couple of applications that foreshadow the heuristic value of thinking of judicial evaluations in terms of academic activity. Two prime examples are the

47 Mark Textor, ‘States of Affairs’ in Edward N Zalta (ed), *The Stanford Encyclopedia of Philosophy* (Summer 2021, Metaphysics Research Lab, Stanford University 2021) <<https://plato.stanford.edu/archives/sum2021/entries/states-of-affairs/>> accessed 23 October 2023.

48 I am subscribing to the broad understanding of “science” as elucidated by Willard Van Orman Quine, *From Stimulus to Science* (Harvard University Press 1998) 49.

49 Uwe Kischel, ‘Tertium Comparationis’ in Jan M Smits and others (eds), *Elgar Encyclopedia of Comparative Law* (Edward Elgar Publishing Limited 2023) <<https://www.elgaronline.com/display/book/9781839105609/b-9781839105609.tertium.comparationis.xml>> accessed 23 October 2023.

50 See eg Thomas F Lüscher, ‘Measuring the Unmeasurable: Assessing the Quality of Science and Scientists’ (2018) 39 *European Heart Journal* 1765.

51 Ilkka Niiniluoto, ‘Scientific Progress’ in Edward N Zalta (ed), *The Stanford Encyclopedia of Philosophy* (Winter 2019, Metaphysics Research Lab, Stanford University 2019) <<https://plato.stanford.edu/archives/win2019/entries/scientific-progress/>> accessed 23 October 2023.

52 Kelley D Mayden, ‘Peer Review: Publication’s Gold Standard’ (2012) 3 *Journal of the Advanced Practitioner in Oncology* 117, 121.

application of two-way anonymity to the evaluation of the cases decided by a judge and the involvement of non-judicial, lay evaluators in the assessment of judicial performances. As for the former innovation, the idea of applying scientific research standards to judicial performance evaluation is not unprecedented. In a key volume edited by Contini, authors Bencze, Kovács, and Zódi contemplated the possibility of introducing an academic-style review process in Hungarian judicial performance evaluations:

For the sake of excluding prejudice and exacting a uniform application of the law, the evaluation of the judicial activity should be carried out on the basis of the system that is used during the quality examination of scientific publications. Therefore, assessment of judgments rendered by a judge could be trusted to professionally renowned fellow justices functioning at other regional courts of law, who would give their opinion on the particular judge's work based on anonymized decisions and case files ('blind peer-review'). This way, disparities of legal practice and reasoning style within the country may be brought to the surface more easily apart from the objective assessment of the particular judge. This kind of blind peer-review system could bring awareness to dispensing justice.⁵³

Despite this exception, however, it is an understatement to say that the idea of importing anonymized peer review mechanisms to judicial performance evaluations has been neglected in the literature. This is an unfortunate circumstance since the implementation of double-blind assessments to judicial performance evaluations arguably addresses the one major limitation plaguing human-based methods: the risk of bias. Of course, such kind of anonymization would require other auxiliary precautions for the system to work: for example, both the parties of the case and the general context should be sufficiently anonymized so that the evaluator cannot discover the identity of the evaluatee by other means.

As per the second-mentioned innovation, to think of judicial performance evaluations in terms of academic scholarship raises the additional question of whether the evaluation of judgments and other products of judicial activity should be carried out only by fellow magistrates. Who is to be, one may ask, the judge's "peer"? To the extent that scholars, lawyers, or other trusted legal experts are institutionally expected to be equally versed in the understanding of the law, they may somehow take part in the evaluation. Surely such participation should be most carefully weighed lest the principle of judicial independence be jeopardized. Nonetheless, it may open up a welcome opportunity for enhancing the external accountability of the judicial branch, which is often marred by accountability deficits.⁵⁴ After all, legal scholars "review" judicial work every time they publish case notes.

⁵³ Mátyás Bencze, Ágnes Kovács and Zsolt Zódi, 'The Evaluation and Development of the Quality of Justice in Hungary' in Francesco Contini (ed), *Handle with Care Assessing and designing methods for the evaluation and development of the quality of justice* (IRSIG-CNR 2017) 151 <https://www.academia.edu/36129389/Handle_with_Care_Assessing_and_designing_methods_for_the_evaluation_and_development_of_the_quality_of_justice> accessed 14 May 2023.

⁵⁴ David Kosař, 'The Least Accountable Branch' (2013) 11 *International Journal of Constitutional Law* 234.

CLOSING REMARKS

Evaluating judicial performances is a daunting task fraught with perils and pitfalls. At the outset, the intellectual enterprise of applying general rules to concrete cases seems a far cry from the kind of standardizable behavior that is normally operationalized in job performance appraisals. Moreover, the complexities of the judicial process complicate the use of the most readily available data, such as reversal rates or disposition time. However, the peculiar features of judicial activity do not make the case for evaluation of the work of judges any less cogent, or feasible. Besides the minor note that an appreciable part of judicial workload can indeed be standardized, a major argument can be lodged as to the fact that other peculiar human enterprises, such as scientific research, have been extensively and widely subjected to quality and ranking measurements despite clear operationalization issues. To put it otherwise, if rankings are possible in the heavily debated world of scientific research, we should not flinch at the task of evaluating and ranking judicial capabilities.

Besides confirming the opportunity to evaluate output quality, the parallel with scientific research “writ large” proves heuristic as to possible solutions. To the extent that judicial activity nears scientific scholarship as a goal-oriented enterprise where quality is hard to assess, to think of the former in terms of the latter may yield beneficial ideas for innovation. Current methods of judicial performance evaluation involve the use of surveys and interviews or hinge on the use of some available “objective” data. However, these traditional avenues to judicial performance assessment fall under scrutiny for the risk of bias, or because they do not provide sufficiently unequivocal meaning to assess the judge’s performance. In this respect, the difficulties faced by current methodologies may be short-circuited through the implementation of a system of double-blind peer review system, which is, as it happens, the shibboleth of modern academic scholarship.

FUNDING INFORMATION

This work was supported in part by the Ministry of Justice program “UNI 4 JUSTICE — Universitas per la Giustizia. Programma per la qualità del sistema giustizia e per l’effettività del giusto processo” (CUP J19J21026980006).

COMPETING INTERESTS

The author has no competing interests to declare.

AUTHOR AFFILIATIONS

Dr. **Silvio Roberto Vinceti**  orcid.org/0000-0002-1536-5745
Research Fellow, University of Modena and Reggio, Italy

REFERENCES

Bencze M, Kovács Á and Zódi Z, ‘The Evaluation and Development of the Quality of Justice in Hungary’ in Francesco Contini (ed), *Handle with Care Assessing and designing methods for the evaluation and development of the quality of justice* (IRSIG-CNR 2017) <https://www.academia.edu/36129389/Handle_with_Care_Assessing_and_designing_methods_for_the_evaluation_and_development_of_the_quality_of_justice> accessed 14 May 2023

- Benvenuti S, 'The Italian System of Judicial Governance: An Arena of Confronting Informal Practices and the Push Towards Formalization' (2023) 24 *German Law Journal* 1373. DOI: <https://doi.org/10.1017/glj.2023.67>
- Berch RW and Bass EN, 'Judicial Performance Review in Arizona: A Critical Assessment' (2014) 4 *Oñati Socio-Legal Series* 927
- Berger R, *Government by Judiciary: The Transformation of the Fourteenth Amendment* (1977) (2nd edn, Liberty Fund 1997)
- Bork RH, *Tempting of America. The Political Seduction of the Law* (Simon & Schuster 1990)
- Burdese A, 'Voce Magistrato (Dir. Rom.)' 187
- Catino M, Dallara C and Rocchi S, 'The Organizational Reasons for Wrongdoing. The Case of Italy's Superior Council of the Judiciary (CSM)' (2023) 79 *Crime, Law and Social Change* 453. DOI: <https://doi.org/10.1007/s10611-022-10064-9>
- Choi S and Gulati M, 'A Tournament of Judges?' (2004) 92 *California Law Review* 299. DOI: <https://doi.org/10.2307/3481449>
- Contini F (ed), *Handle with Care Assessing and Designing Methods for the Evaluation and Development of the Quality of Justice* (IRSIG-CNR 2017) <https://www.academia.edu/36129389/Handle_with_Care_Assessing_and_designing_methods_for_the_evaluation_and_development_of_the_quality_of_justice> accessed 14 May 2023
- Contini F, Mohr R and Velicogna M, 'Formula over Function? From Algorithms to Values in Judicial Evaluation' (2014) 4 *Oñati Socio-Legal Series* 1099
- Di Federico G (ed), *Recruitment, Professional Evaluation and Career of Judges and Prosecutors in Europe: Austria, France, Germany, Italy, the Netherlands and Spain* (Lo scarabeo 2005)
- , 'Magistrati ordinari: Organico, Reclutamento, Valutazioni della professionalità, Carriera, Tramutamenti e attribuzione delle funzioni giudiziarie, Trattamento economico' in Giuseppe Di Federico (ed), *Ordinamento giudiziario: uffici giudiziari, CSM e governo della magistratura* (Bononia University Press 2019)
- , 'Riforma delle valutazioni di professionalità e delle disfunzioni ad essa connesse, in prospettiva comparata' [2022] *Diritto di difesa* 1
- Elek J and Rottman D, 'Methodologies for Measuring Judicial Performance: The Problem of Bias' (2014) 4 *Oñati Socio-Legal Series* 863
- Elek JK, 'Judicial Perspectives on Emotion, Emotion Management, and Judicial Excellence in the USA' (2019) 9 *Oñati Socio-Legal Series* 865. DOI: <https://doi.org/10.35295/osls.iisl/0000-0000-0000-1033>
- Fabri M, 'Clash of Visions: Regulating Judges and Prosecutors in Italy', *Regulating Judges* (Edward Elgar Publishing 2016) <<https://www.elgaronline.com/edcollchap/edcoll/9781786430786/9781786430786.00018.xml>> accessed 22 April 2024. DOI: <https://doi.org/10.4337/9781786430793.00018>
- Gill RD, 'Implicit Bias in Judicial Performance Evaluations: We Must Do Better Than This' (2014) 35 *Justice System Journal* 301. DOI: <https://doi.org/10.1080/0098261X.2013.873290>
- Gill RD, Lazos SR and Waters MM, 'Are Judicial Performance Evaluations Fair to Women and Minorities? A Cautionary Tale from Clark County, Nevada' (2011) 45 *Law & Society Review* 731. DOI: <https://doi.org/10.1111/j.1540-5893.2011.00449.x>
- Hagsgård MB, 'Internal and External Dialogue: A Swedish Approach to Quality Work in Courts' (2014) 4 *Oñati Socio-Legal Series* 993
- Hamřík L, 'Actors of Informal Judicial Institutions and Practices' (2023) 24 *German Law Journal* 1520. DOI: <https://doi.org/10.1017/glj.2023.91>
- Hurst W, 'Who Is the "Great" Appellate Judge?' (1949) 24 *Indiana Law Journal* 394

- Kearney RC, 'Judicial Performance Evaluation in the States' (1999) 22 Public Administration Quarterly 468
- Kischel U, 'Tertium Comparationis' in Jan M Smits and others (eds), *Elgar Encyclopedia of Comparative Law* (Edward Elgar Publishing Limited 2023) <<https://www.elgaronline.com/display/book/9781839105609/b-9781839105609.tertium.comparationis.xml>> accessed 23 October 2023. DOI: <https://doi.org/10.4337/9781839105609.tertium.comparationis>
- Kosař D, 'The Least Accountable Branch' (2013) 11 International Journal of Constitutional Law 234. DOI: <https://doi.org/10.1093/icon/mos056>
- Lawson G, 'Equivocal Originalism' (2022) 27 Texas Review of Law and Politics 309. DOI: <https://doi.org/10.2139/ssrn.4050058>
- Leiter B, 'Legal Indeterminacy' (1995) 1 Legal Theory 481. DOI: <https://doi.org/10.1017/S1352325200000227>
- Lüscher TF, 'Measuring the Unmeasurable: Assessing the Quality of Science and Scientists' (2018) 39 European Heart Journal 1765. DOI: <https://doi.org/10.1093/eurheartj/ehy295>
- Mangione G, 'Some Brief Remarks on the Controversial Relationship Between the Judiciary and Politics in Italy' (2021) 27 Comparative Law Review 79. DOI: <https://doi.org/10.12775/CLR.2021.003>
- Mayden KD, 'Peer Review: Publication's Gold Standard' (2012) 3 Journal of the Advanced Practitioner in Oncology 117. DOI: <https://doi.org/10.6004/jadpro.2012.3.2.8>
- McIntyre J, 'Evaluating Judicial Performance Evaluation: A Conceptual Analysis' (2014) 4 Oñati Socio-Legal Series 898
- Melville A, 'Evaluating Judicial Performance and Addressing Gender Bias' (2014) 4 Oñati Socio-Legal Series 880
- Mitchell G, 'Evaluating Judges' in David E Klein and Gregory Mitchell (eds), *The Psychology of Judicial Decision Making* (Oxford University Press 2010) accessed 15 May 2023. DOI: <https://doi.org/10.1093/acprof:oso/9780195367584.003.0014>
- Musolf L, 'Performance Evaluation of Federal Administrative Law Judges: Challenge for Public Administration?' (1998) 28 The American Review of Public Administration 390. DOI: <https://doi.org/10.1177/027507409802800404>
- Niiniluoto I, 'Scientific Progress' in Edward N Zalta (ed), *The Stanford Encyclopedia of Philosophy* (Winter 2019, Metaphysics Research Lab, Stanford University 2019) <<https://plato.stanford.edu/archives/win2019/entries/scientific-progress/>> accessed 23 October 2023
- OECD, *Government at a Glance 2021* (Organisation for Economic Co-operation and Development 2021) <https://www.oecd-ilibrary.org/governance/government-at-a-glance-2021_1c258f55-en> accessed 22 October 2023
- Olanescu A, 'Discretionary Power and Abuse of Power in the Activity of State Institutions' (2021) 84 Curentul Juridic 80
- Pauliat H and others, 'The Evaluation and Development of the Quality of Justice in France' in Francesco Contini (ed), *Handle with Care Assessing and designing methods for the evaluation and development of the quality of justice* (IRSIG-CNR 2017) <https://www.academia.edu/36129389/Handle_with_Care_Assessing_and_designing_methods_for_the_evaluation_and_development_of_the_quality_of_justice> accessed 14 May 2023
- Posner EA, 'The Constitution of the Roman Republic' in Giuseppe Dari-Mattiacci and Dennis P Kehoe (eds), *Roman Law and Economics: Institutions and Organizations Volume I* (Oxford University Press 2020) <<https://doi.org/10.1093/oso/9780198787204.003.0003>> accessed 14 May 2023

- Preshova D, Damjanovski I and Nechev Z, 'The Effectiveness of the "European Model" of Judicial Independence in the Western Balkans: Judicial Councils as a Solution or a New Cause of Concern for Judicial Reforms' [2017] *Cleer Papers* 7
- Przybyszewski L, 'The Dilemma of Judicial Biography or Who Cares Who Is the Great Appellate Judge? Gerald Gunther on Learned Hand' (1996) 21 *Law & Social Inquiry* 135. DOI: <https://doi.org/10.1086/492542>
- Quine WVO, *From Stimulus to Science* (Harvard University Press 1998)
- Reddick M and Love Kourlis R, 'Judicial Performance Evaluation' in Peter Koelling and Rebecca Love Kourlis (eds), *The improvement of the administration of justice* (American Bar Association 2016)
- Singer JM, 'Judicial Performance Evaluation in the States: The IAALS JPE 2.0 Pre-Convening White Paper' [2022] *SSRN Electronic Journal* <<https://www.ssrn.com/abstract=4505778>> accessed 22 October 2023. DOI: <https://doi.org/10.2139/ssrn.4505778>
- Solum LB, 'A Tournament of Virtue Symposium: Empirical Measures of Judicial Performance' (2005) 32 *Florida State University Law Review* 1365. DOI: <https://doi.org/10.2139/ssrn.588322>
- , 'Originalist Methodology' (2017) 84 *University of Chicago Law Review* 269
- Stella A, 'Intervista al professor Di Federico: "Grazie a Palamara hanno fatto carriera anche i giudici che lo giudicheranno"' *Il Riformista* (22 July 2020) <<https://www.ilriformista.it/intervista-al-professor-di-federico-grazie-a-palamara-hanno-fatto-carriera-anche-i-giudici-che-lo-giudicheranno-134353/>> accessed 14 May 2023
- Stella V, 'Intervista a Eugenio Albamonte. «Valutare Noi Toghe? No a Un'élite Di Eccellenti, Ma i Capi Vanno Formativi»' *Il Dubbio* (24 March 2021) 1
- Stiers D, 'Performance Voting, Retrospective Voting, and Economic Voting. Conceptual Clarity and Empirical Testing' (2022) 103 *Social Science Quarterly* 399. DOI: <https://doi.org/10.1111/ssqu.13125>
- Tate CN and Vallinder T (eds), *The Global Expansion of Judicial Power* (New York University Press 1995)
- Textor M, 'States of Affairs' in Edward N Zalta (ed), *The Stanford Encyclopedia of Philosophy* (Summer 2021, Metaphysics Research Lab, Stanford University 2021) <<https://plato.stanford.edu/archives/sum2021/entries/states-of-affairs/>> accessed 23 October 2023
- Tsereteli N, 'Constructing the Pyramid of Influence: Informal Institutions as Building Blocks of Judicial Oligarchy in Georgia' (2023) 24 *German Law Journal* 1469. DOI: <https://doi.org/10.1017/glj.2023.74>
- Van Dijk D and Schodl MM, 'Performance Appraisal and Evaluation' in James D Wright (ed), *International Encyclopedia of the Social & Behavioral Sciences (Second Edition)* (Elsevier 2015) <<https://www.sciencedirect.com/science/article/pii/B9780080970868220345>> accessed 12 May 2023. DOI: <https://doi.org/10.1016/B978-0-08-097086-8.22034-5>
- Vinceti SR, 'Standardisation and Authority in Judicial Discipline: A Comment on Italy and Ireland's Experiences' (2022) 6 *Irish Judicial Studies Journal* 55
- Vincze A, 'Schrödinger's Judiciary—Formality at the Service of Informality in Hungary' (2023) 24 *German Law Journal* 1432. DOI: <https://doi.org/10.1017/glj.2023.71>
- Warren M, 'From Evaluation to Improvement: A Chief Justice's Perspective' (2014) 4 *Oñati Socio-Legal Series* 953

Vinceti
*International Journal
 for Court Administration*
 DOI: 10.36745/ijca.576

TO CITE THIS ARTICLE:

Silvio Roberto Vinceti,
 'Innovating Judicial
 Performance Evaluations:
 Toward Academic-Style
 Peer Review?' (2024) 15(1)
*International Journal
 for Court Administration*
 2. DOI: [https://doi.
 org/10.36745/ijca.576](https://doi.org/10.36745/ijca.576)

Published: 10 June 2024

COPYRIGHT:

© 2024 The Author(s).
 This is an open-access
 article distributed under
 the terms of the Creative
 Commons Attribution
 4.0 International License
 (CC-BY 4.0), which
 permits unrestricted
 use, distribution, and
 reproduction in any
 medium, provided the
 original author and source
 are credited. See [http://
 creativecommons.org/
 licenses/by/4.0/](http://creativecommons.org/licenses/by/4.0/).

*International Journal for
 Court Administration* is
 a peer-reviewed open
 access journal published by
 International Association
 for Court Administration.

