

Evaluating Judicial Performance Evaluation: A Conceptual Analysis

JOE MCINTYRE*

McIntyre, J., 2014. Evaluating Judicial Performance Evaluation: *A Conceptual Analysis*. *Oñati Socio-legal Series* [online], 4 (5), 898-926. Available from: <http://ssrn.com/abstract=2533854>



Abstract

While the performance evaluation of judges has become a ubiquitous aspect of modern judicial administration, evaluation mechanisms are too often utilised uncritically, without reflection on *why* we evaluate judges, and how 'measurement' furthers these objectives. This article provides a conceptual analysis of the role and purpose of performance evaluation, conceiving it as a limited tool of judicial accountability, which itself exists only to promote excellent judging. As such, the efficacy of evaluation mechanisms must always be assessed by reference to their impact on these overarching accountability objectives. The article explores the value of this conception approach by briefly examining three uses of performance evaluation: 1) judicial promotions; 2) judicial retention elections; and 3) judicial professional development. In doing so it illustrates how a clear conceptual approach invites a more nuanced and critical examination of the limitations and benefits of judicial performance evaluation programs.

Key words

Judicial performance evaluation; judicial theory; judges; judicial method; judicial retention election; judicial promotion

Resumen

Mientras que la evaluación del rendimiento de los jueces se ha convertido en un aspecto omnipresente de la administración judicial moderna, los mecanismos de evaluación se utilizan con demasiada frecuencia de manera acrítica, sin reflexionar sobre las razones por las que evaluamos a los jueces, y cómo se alcanzan los objetivos buscados mediante la 'medición'. Este artículo ofrece un análisis conceptual de la función y el propósito de la evaluación del rendimiento, concibiéndola como una herramienta limitada de la responsabilidad judicial, que a

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).

* Dr Joe McIntyre is a Lecturer in Law at Charles Darwin University, Australia. Dr McIntyre is a judicial theorist who has taught in Australia, Canada and the UK. He obtained his PhD from the University of Cambridge, where his thesis examined the nature of the judicial function. He can be contacted at: School of Law, Charles Darwin University, Ellengowan Drive, Darwin, Northern Territory, Australia; or by email at joe.mcintyre@cdu.edu.au



su vez sólo existe para promover la excelencia judicial. Como tal, la eficacia de los mecanismos de evaluación siempre se debe comprobar en función de su impacto en los objetivos de responsabilidad globales. El artículo explora el valor de este enfoque examinando brevemente tres usos de la evaluación de rendimiento: 1) promociones judiciales; 2) elecciones para la reelección de jueces; y 3) el desarrollo profesional judicial. Al hacerlo, se refleja cómo un enfoque conceptual claro invita a un examen más matizado y crítico de las limitaciones y beneficios de los programas de evaluación del rendimiento judicial.

Palabras clave

Evaluación del rendimiento judicial; teoría judicial; jueces; método judicial; elecciones a la reelección de jueces; promoción judicial

Table of contents

1. Introduction.....	901
2. The origins of judicial performance evaluation.....	902
2.1. Different justifications and purposes	904
3. The concept of judicial accountability.....	905
3.1. The nature of judicial accountability.....	905
3.2. The judicial function and integrity	906
3.3. The 'external objective' and 'internal subjective' aspects of accountability	907
3.4. A limited conception of judicial accountability	907
4. Judicial performance as a tool of judicial accountability	908
5. Examining the implications of the conception	909
5.1. Performance evaluation and judicial promotions	910
5.1.1. Performance evaluation and the problem of proxies.....	910
5.2. Performance evaluation informing judicial retention elections	912
5.2.1. Performance evaluation and the provision of information to voters	912
5.2.2. Criticism of evaluation programs.....	913
1. Inaccurate assessments	913
2. Limited use of information.....	914
3. Potential for undue influence	915
5.2.3. Critique of the primary mechanism	915
5.3. Performance evaluation and judicial professional development.....	917
5.3.1. The role of performance evaluation in informing professional development.....	917
5.3.2. Limitations and criticisms	918
6. Conclusions	919
6.1. A Conceptual framework for analysing judicial performance evaluation programs	919
6.1.1. application of framework	920
6.2. Final observations on evaluation programs	921
References	921

1. Introduction

The evaluation of judicial performance of judges is, in one sense, a deeply familiar and traditional aspect of the practice of law. The lawyer advising the client whether to appeal, the academic analysing the latest Supreme Court decision and even, rarely, the Parliament considering the potential disciplining of a judge, are all engaged in a form of *evaluating* judicial performance. To the modern judicial scholar, however, the term 'judicial performance evaluation' has taken on a new and highly particularised usage, referring to a distinct and discrete tool of judicial administration: the systematic survey and measurement of judicial performance. This accountability tool has, intriguingly, emerged as something of a juridical universal. The language of 'judicial performance evaluation' is familiar to judicial scholars across common law and civil law jurisdictions, as evidenced by the workshop from which this journal volume evolved. Despite the diverse judicial milieu from which participants were drawn, it was largely taken for granted that a conversation on the modern practice and implications of judicial performance evaluation was possible, and indeed unremarkable. The language of 'judicial performance evaluation' was taken to convey a discrete and definite meaning, referring to a distinct tool of judicial administration.¹

This new tool has become popular with judicial administrators and politicians, keen to ensure that judicial officers and institutions achieve administrative efficiency and are properly held to account. However, the use of this tool is often highly controversial - critics have concern over the potential of the tool to undermine judicial independence and impartiality, distorting decision-making. The judiciary has been hesitant about, if not outright hostile to, any systemic approach to assess or evaluate judges (See Brennan *et al.* 1986, p. 77).² This controversial nature is evident in the extraordinary cautionary tale of Riddel *et al.* (2012) who were forced to twice abandon well-conceived empirical studies into judicial performance evaluation following a sustained outcry and refusal to cooperate from both the judiciary and the profession. There is legitimate concern that systematic evaluations may become a tool for executive interference (Colbran 2002b, p. 248). Judiciaries around the world are increasingly being placed under mounting pressure, through tightening budget, emboldened politicians and media increasingly willing to attack the judiciary (Brandenburg 2009, p. 371).³ This has been exacerbated by a public clamour for more judicial accountability in many jurisdictions (Feltner 2008, p. 177)⁴ and by academic critique of traditional accountability mechanisms (See Colbran 2003b). In such a context it should not be surprising that there will be resistance to *any* proposed new form of accountability. Judicial performance evaluation is not, however, simply another form of accountability, but rather has become the focal point for many of these debates. Proponents argue that judicial performance evaluation is a veritable wonder-drug, allowing meaningful accountability while protecting judicial independence (See Sponzo 1987, Brody 2000, 2008, White 2002, 2009, Colbran 2002b, Kourlis and Singer 2007, Haines 2010). It is seen as: a valuable aid to judicial professional development; creating more open judiciaries; making courts more financially transparent; motivating judges to be more productive; enhancing democratic engagement; and providing objective means of promoting judges (Colbran 2003b, Warren 2011, Riedel 2014, Berch and Bass 2014). Critics, however, argue that judicial performance evaluation threatens judicial independence and creates

¹ For a useful discussion of the concept and context of judicial performance evaluation, see the article of Roach Anleu and Mack (2014, p. 1018-1020) in this volume.

² This hostility has been particularly evident in the common law world, though as Mohr and Contini (2007, p. 10, 23) note, it is clearly evident in civil systems as well.

³ For a description of some of the extraordinary and outrageous attempts to intimidate judges in the last 20 years in the USA, see Brandenburg (2009, p. 377-379).

⁴ Tellingly, Feltner observes that this clamour arises despite that fact that the 'public sometimes does not seem to be quite sure of what kind of accountability they mean, or of what precise problems require more accountability' (Feltner 2008, p. 177).

opportunities for undue executive interference; undermines impartiality and distorts decision-making; misconceives the role of the judiciary; improperly favours measurable indicators over evaluative criteria; it improperly imports fashions of managerialism; is biased against minorities; is expensive, time-consuming and ultimately counter-productive (Griffin 1995, Mohr *et al.* 1996, Drummond 2001a, 2001b, Spigelman 2002, 2006, Contini and Mohr 2007, Rares 2011, Elek *et al.* 2012, Gill 2012,).

The appropriateness and proper scope of judicial performance evaluation is inherently contested ground. This is particularly so given that while the mechanisms of judicial performance evaluation are now deeply familiar to judicial studies scholars, very little is written about precisely what the concept captures, or what are its proper purposes or limits. The lack of conceptual analysis of judicial performance evaluation has meant that critics and proponents often seem to be talking at cross purposes. This article addresses this conceptual deficit by providing a theoretical framework⁵ to understand the purpose, scope and limits of judicial performance evaluation. Rather than focusing on the qualities that performance evaluation can measure, I address the anterior issues of why measure at all. I argue that the term 'judicial performance evaluation' has come to represent a particular form of evaluation that relies upon quantitative measures to provide 'objective' analysis of judicial performance. That analysis is largely used to support related mechanisms of judicial accountability. In this paper I examine three particular usages, namely the use of judicial performance evaluation as a tool of judicial professional self-improvement, as an aid to judicial retention elections, and as an aid to judicial promotions.⁶ I conclude by arguing judicial performance evaluation has a limited and derivative usage. It must always be evaluated by reference to its ability to promote the ends of judicial accountability, with the accountability benefits outweighing its – broadly assessed – costs.

2. The origins of judicial performance evaluation

To understand the modern concept of judicial performance evaluation it is useful, like for all things, to start at the beginning. Like all contested things, however, that beginning is difficult to identify. There are several threads that run through the modern usage of the term, each with a distinct origin.⁷ The *language* of judicial performance evaluation, at least, first began to appear in the US in the late 1970s. Certainly, the first official state-sponsored 'judicial performance evaluation' program began in Alaska in 1976 (See Brennan *et al.* 1986, p. 78-82, Elek *et al.* 2012, p. 65)⁸ as part of an effort provide better information for retention elections (Brody 2008, p. 118, Elek, *et al.* 2012, p. 65). In contrast to the pre-existing and problematic programs of 'Bar Polls' (See Aynes 1981, p. 268, Pelander 1998, p. 648, White 2002, p. 1064-1066), the new approach sought to provide more reliable and accurate information for voters. The concept of judicial performance evaluation was picked up by an internal paper of the National Center on State Courts,⁹ and the ABA¹⁰ where the focus shifted away from evaluation for retention purposes towards professional development (Aynes 1981, p. 272). Even at this early stage, this divergence of rationale affected the design of appropriate mechanisms

⁵ I note that when I use the term 'theory' in this paper in the sense of 'legal theory', that is, the process of providing clear analytic analysis of concepts, institutional functions and role. A do not use the term as referring to 'a theory' in the positivistic assertion sense of a position demonstrable through empirical analysis.

⁶ I explore these issues by reference to the articles of Berch and Bass (2014) and Riedel (2014) published in this volume, and the work of another participant, Warren (2011).

⁷ It is interesting to note that Aynes, writing at the beginning of the emergence of the judicial performance evaluation movement already recognised that the 'impetus for such a program of judicial evaluation is difficult to trace' (Aynes 1981, p. 261).

⁸ It is worth noting that the Alaskan approach was followed in 1978 by New Jersey (Aynes 1981, p. 261).

⁹ See National Center on State Courts, '*Judicial Performance Evaluation: Concept Paper*', (Draft 1979).

¹⁰ American Bar Association, '*Evaluation of Judicial Performance Committee: Concept Paper*', (1979). See Aynes (1981, p. 263).

(See Aynes 1981, p. 269). By the beginning of the 1980s there was a clear sense that use of judicial performance evaluation to improve the quality of judicial work was an idea whose time had come.¹¹

This rapid emergence reflected the seductive power behind the name of this new tool. The management tool of job performance evaluation had been gaining traction as a tool of management in private industry, the public sector, and education.¹² The language of 'performance evaluation' was the fashion of the moment in management, and quickly spread to this new judicial context (Spigelman 2002, p. 20). In doing so it perhaps unwittingly imported into the judicial context management ideas together with management language. This rapidly expanded with the emergence of theories of 'new public management', which emphasised values of efficiency and a private-enterprise mentality (See Ng 2007, p. 11).¹³ This new public management approach transformed and invigorated the use of judicial performance evaluation. 'Evaluation' transformed from a qualitative to a quantitative enterprise, striving to measure 'objective' outputs in a way that was apparently free from bias and improper subjectivity. This attractive new form sought to introduce 'managerial' systems of judicial accountability that conceived of the judicial system in terms of outputs (Mohr and Contini 2007, p. 10, 19). The traditional judicial forms of *legal accountability* were overlaid by new methods of *managerial accountability* that strove to 'protect and promote efficiency, cost control, and link results to resources' (Contini and Mohr 2007, p. 30). This movement was attractive to many as it seems self-evident that citizens are entitled to know that public funds are being spent efficiently and effectively (Spigelman 2002, p. 19). This model appeared to present a means of increasing judicial accountability, increasing performance, and enhancing efficiency without interfering with judicial independence (See White 2002, Colbran 2002b, Brody 2008, Haines 2010). However, from the beginning the managerialisation of judicial accountability met with resistance, as the judge was no longer valued exclusively as an independent decision-maker, but also as an actor in a 'public organisation delivering services to the public' (Contini and Mohr 2007, p. 27-8). This represented a revolutionary re-conception of the judicial role (See Spigelman 2000b, p. 380, 2002, p. 26). Management objectives of productivity were seen to diverge and conflict with judicial objectives of impartiality and independence (Contini and Mohr 2007, p. 26). It became feared that judicial accountability would come to be seen 'exclusively from a managerial perspective' (Contini and Mohr 2007, p. 28), where all that mattered would be those things that could be counted and measured (Spigelman 2006, p. 69). Management practices gave a name to a new tool of judicial accountability, and have come, over time to heavily shape the modern practice of judicial performance evaluation. Judicial performance evaluation cannot be divorced from the managerialist context. There are clear originating forces located in ideas of judicial self-improvement and public democratic engagement, though these do not tell the whole story. The movement towards measurable and quantifiable indicators of judicial standards arose in the context of arguments about economic efficiency. This association continues to create hostility towards the tool. To avoid such hostility it is necessary to clearly articulate the motivation and purposes for which the tool of judicial performance evaluation will be used.

¹¹ Aynes suggests the idea 'probably came to many people across the country almost simultaneously' (Aynes 1981, p. 261).

¹² Aynes describes this model as operating 'on a very basic premise of improving human behavior: that desirable conduct should receive positive reinforcement and that areas for improvement should be identified so that the individual in question can improve performance' (Aynes 1981, p. 262).

¹³ Spigelman (2002, p. 19) has described the dominant characteristic of the new managerialist focus as 'a wish to replicate a results-driven model, said to be characteristic of the private sector where the free operation of market forces ensures efficient use of resources. This approach is said to require the explicit identification of goals, which must be measurable so that performance can be assessed in what is regarded as an "objective" manner'.

2.1. Different justifications and purposes

Given the history from which judicial performance evaluation arose, it is unsurprisingly difficult to articulate *the* purpose of the tool. Each of the different originating influences – ideas of selection and retention of judges, of professional improvement, and of measurable efficiency – have left indelible marks on the modern form of judicial performance evaluation.

While there seems to be broad recognition that judicial performance evaluation programs help courts achieve a variety of goals, articulating these goals are is not always easy. In the US the emphasis has often been on informing voters in retention elections,¹⁴ though other goals such as including improving judicial quality and aiding transparency have also been identified (Elek *et al.* 2012, p. 65). Kourlis and Singer (2008, p. 8-9) argue that such programs serve three objectives:

- 1) 'to provide constructive feedback to sitting judges to inform their professional development;
- 2) to educate the public on the work of its judges and foster appropriate expectations about the role of the judge; and
- 3) where applicable, to provide relevant information to decision-makers concerning the retention or reappointment of judges.'¹⁵

The educative purpose introduces a new strand into the discourse, recognising that the mere use of a tool such as judicial performance evaluation will create ripples in society: it may educate the public, or it may serve as a reminder to the judiciary that the public are entitled to superior performance from the judiciary (Colbran 2003b, p. 68). These ripples may be desirable, but there can be unintended consequences. Some of these consequences have played a strong role in shaping the use of judicial performance evaluation programs in Europe (See Contini and Mohr 2007, p. 36, 39). The career judiciaries of the continental civil legal systems provide a different set of demands on the administration of justice, and in this context the management and managerialist aspects of judicial performance evaluation have come to the fore. The evaluation of judges is seen as a key 'tool of staff management' (Riedel 2014, p. 977), though often it has also been seen as a tool of economic efficiency (Contini and Mohr 2007, p. 37). In Australia, programs of judicial performance evaluation have been more limited, with the focus mostly on court and administrative performance measurement¹⁶ as part of a strategic approach to judicial institutional management (Mohr *et al.* 1996, p. 157). In more recent years the systematised tools of performance evaluation have, however, been also utilised for judicial professional development (See Warren 2011).

The different purposes for which judicial performance evaluation has been utilised throughout the world make it difficult isolate a common theme unifying these practices. While there is the real possibility that it is an illusion to talk of a global practice of 'judicial performance evaluation' – that there is only a collection of unrelated practices that coincidentally share the same name – there does seem to be a common underlying objective that unifies these different usages. These different programs all seek to utilise structured, tightly-regulated and repeatable methodologies to *measure* judicial performance. By the use of standardised interview and questionnaires, analysis of outputs and performance indicators, and similar techniques, programs of judicial performance evaluation strive to provide

¹⁴ Griffin, for example, argues that beneath the various schemes there is constant underlying premise that such programs operate to 'enhance 'voter awareness' and popular participation in judicial elections' (Griffin 1995, p. 5). Brody has expanded on this electoral purpose, arguing that informed voting operates to 'thereby ensure a high-quality judiciary' (Brody 2000, p. 334) though does not articulate the precise mechanism by which this may occur.

¹⁵ It is worth noting that the third of these objectives is phrased in such a way as to capture reappointment mechanisms other than retention elections.

¹⁶ This has been described as a focus on the 'time and motion' of judicial activity (Colbran 2002b, p. 236).

feedback on a judge's performance in a way that does not depend upon mere opinion of a few (often partial) individuals.¹⁷ The use of that feedback is directed, through a number of different paths, towards a common purpose: the promotion of judicial accountability. To understand the modern practice of judicial performance evaluation (and its limitations) it is necessary to understand its relationship to the concept of judicial accountability.

3. The concept of judicial accountability

While society places 'enormous trust and confidence' (Gleeson 1979, p. 338)¹⁸ in the judiciary, trust alone is insufficient. We insist upon mechanisms of 'judicial accountability' to promote the integrity, excellence and responsiveness of the judiciary. The basic imperative for accountability may be irresistible:¹⁹ 'judicial accountability' appears to be a 'self-evident good' (Handsley 2001b, p. 181), something that 'everyone is 'for', like democracy or freedom' (Spigelman 2002, p. 18). However, the realisation of this imperative is problematic; the scope of the concept is unclear and the literature lacks theoretical clarity. This conceptual imprecision leaves the concept liable to be co-opted and misused (Geyh 2008, p. 912), allowing the 'drumbeat of judicial accountability' to drown out other judicial values.²⁰ As Sandra Day O'Connor (2008, p. 1) notes:

'Judicial accountability...is a concept that is frequently misunderstood at best and abused at worst. It has become a rallying cry for those who want in reality to dictate substantive judicial outcomes.'

This concern is particularly acute in the context of judicial performance evaluation, with fear that such programs present an opportunity for those who wish to improperly influence judicial conduct to clothe their actions in the legitimising pursuit of 'accountability.' This concern is exacerbated because the uncertainty of the proper nature of judicial accountability and its relationship to judicial performance evaluation.

3.1. The nature of judicial accountability

The term 'accountability' is itself an 'amorphous' (Le Sueur 2004, p. 73) and beset with ambiguities: for example, it has some close (Handsley 2001b, p. 180),²¹ but ultimately unclear relationship to the concept of responsibility (See Handsley 2001a, p. 68, Pimentel 2009, p. 14). In the judicial context, the term has evolved away from any simple 'command-and-control' conception, where a third-party has a power of direction (Le Sueur 2004, p. 73). Judicial accountability is *not* about ceding judicial authority to anyone. A judge may be required to 'give account' (a professional 'explanatory' usage) for conduct, but not to derogate responsibility to act. What then is the purpose of judicial accountability? It is not sufficient to examine only the old question 'accountability to whom and for what' (Spigelman 2002, p. 18, see also White 2002, p. 1060-1061). Rather, it is critical we ask: *to what end* we wish to hold judges to account?

Judicial accountability must exist for the limited purpose of ensuring that the judge strives for the excellent performance of the judicial function: to ensure that judges 'do the job they are supposed to do in the way they are supposed to do it' (Handsley 2001b, p. 218). The mechanisms of judicial accountability operate to

¹⁷ Anderson characterises judicial retention evaluation programs as having some common elements: official status; broad-based survey mechanisms; wide dissemination of survey results and recommendations (Anderson 2001, p. 1376).

¹⁸ As Jayawickrama (2002, p. 1091) notes, judicial power is 'power is given on trust'.

¹⁹ As O'Connor notes, judicial accountability is seen as 'a fundamental democratic requirement' (O'Connor 2008, p. 1).

²⁰ Drummond (2001a, p. 304-305) notes that this has had the effect of making any challenge to accountability appear heretical.

²¹ For example, White (2002, p. 1060) argues that accountability simply 'means answerability or responsibility.' See also Kirby 2003, p. 42.

promote the optimal performance of the judicial function, motivating the judge to adhere to the judicial decision-making method, maintain impartiality, avoid the abuse of office, and strive for excellence. Judicial accountability is a limited functional/instrumental concept (Bandes 2006, p. 947, Geyh 2008, p. 916) that promotes the attainment of judicial excellence.

This conception demands, of course, that we articulate what judicial 'excellence' (Gething 2008a, p. 243). One approach is to simply list the characteristics or traits of the 'ideal' judge.²² There are many notable enunciations of such values,²³ yet there is very little sustained analysis of why these particular values matter. An instrumentalist conception of judicial accountability challenges us to look beyond discrete values, to articulate why the pursuit of such values furthers the ends of judicial accountability. This requires an understanding of the nature and purpose of the judicial function.

3.2. *The judicial function and integrity*

It is beyond the scope of this article to provide an extensive examination of the nature of the judicial function.²⁴ It suffices to note that the judicial function has two core aspects: dispute-resolution and social (normative) governance. Judicial decisions are not only a particular type of institutionalised third-party merit-based resolution, but are acts of normative governance - each judicial decision has an impact upon the legal norms it applies. The effects of each decision radiate from beyond the particular dispute, vitalising, clarifying and developing the law. This dual role, dispute resolution and social governance, demand the pursuit of different values by the judge. Dispute-resolution demands finality and a focus upon the individual litigants; Governance demands the pursuit of responsive correctness, focusing on broader social interests and the generalised maintenance of legal norms.

Both roles demand, though, a reputation for judicial integrity – such a reputation provides the institutional legitimacy upon which the judicial function depends. If this reputation for integrity falters, the ability of a judge to perform the judicial function collapses. A judge who acts with blatant partiality or arbitrariness will be ineffective at finally disposing of the dispute: disputants are unlikely to consider themselves bound by a decision that fundamentally abandons the judicial paradigm. Moreover, where there is a general loss of faith in judicial integrity the use of state enforcement mechanisms to ensure compliance will become socially unacceptable.²⁵ Similarly, such an institution without a reputation for integrity cannot perform its social governance role. Judicial governance is legitimised through a discursive process, gaining its normative strength through its persuasiveness, broad acceptance, and ability to promote responsive stability. This demands judges maintain a reputation for fairness, integrity, courageous impartiality, wisdom and excellence; in turn this grants the judicial institution a necessary legitimacy. However, where that reputation fails and the institution is regarded as arbitrary, capricious or unduly aloof, it no longer becomes rational for the individual to be guided by judicial statements as they lack any expectation of consistency. Without integrity, and the reputation for it, the judiciary cannot perform its dispute-resolution task and is incapable of performing its governance function.

²² One illustrative example is provided by Goodman (1982, p. 113-114) who argues that the ideal type of a good judge would display: (1) Neutrality to the parties; (2) Fair mindedness; (3) Being well versed in the law; (4) Ability to write lucidly and logically; (5) Personal integrity; (6) Good physical and mental health; (7) Judicial temperament; and (8) Ability to handle judicial power sensibly.

²³ Similar lists are produced by the ABA (2007); or the appointment criteria of the UK Judicial Appointment Commission (s.d.).

²⁴ For an extensive discussion of this topic, and its impact upon issues of judicial theory and practice, see McIntyre (2013).

²⁵ While an institution sustained by state force *may* be able to exercise a degree of social control, such a 'might-based' institution abandons any pretension of 'judicial' resolution.

To ensure the excellent performance of the judicial function, judicial institutions require mechanisms to ensure the constraints of the judicial method and the standards of impartiality are adhered to. The mechanisms of judicial accountability serve this purpose, promoting both the authentic *actuality* of judicial integrity, and the *reputation* and *appearance* of it. In doing so it creates the institutional legitimacy necessary for the judicial function. Conceiving of judicial accountability in this way reveals both the 'internal' and 'external' aspects of accountability.

3.3. *The 'external objective' and 'internal subjective' aspects of accountability*

Judicial accountability takes on a two-fold nature, promoting the judicial function by maintaining both the actuality of, and reputation for, integrity. These 'internal' and 'external' elements of accountability respond to different aspects of that concept, and justify different mechanisms.²⁶ The internal 'subjective' or 'personal' aspect of judicial accountability is directed towards the individual judge, developing a personal and professional imperative to actually 'do the right thing'. This *personal* accountability comes from the judge's own internal moral compass and personal integrity, not from any 'vulnerability to discipline or other retribution for misdeeds' (Pimentel 2009, p. 16). It rests upon an internalised judicial desire to *actually* adhere to judicial method and pursue excellence because it is the right thing to do. In contrast, the external 'objective' or 'structural' aspect of judicial accountability is directed to the creation and maintenance of an institutional *reputation* for integrity. This distinct element complements the *actual* integrity of personal accountability, ensuring that judges both *act* with integrity and *appear* to do so.²⁷ The institutional reputation for integrity, quality and impartiality is critical to found the social legitimacy upon which both the dispute-resolution and social governance aspects of the judicial function depends.²⁸ Mechanisms of judicial accountability may promote the excellent performance of the judicial function by furthering either or both of the internal and external aspects of accountability.²⁹

3.4. *A limited conception of judicial accountability*

The instrumental nature of this conception of judicial accountability means that while every mechanism may directly promote either internal or external aspects of judicial accountability, it must ultimately be assessed by reference to its impact on the performance of the judicial function.³⁰ Judicial accountability *cannot* operate for its own ends: a non-instrumental conception of judicial accountability would distort judicial decision-making and undermine the very impartiality, integrity and confidence it should promote.³¹ Judicial accountability possesses an inherently limited and constrained nature: every judicial accountability mechanism must operate in a 'way that does not damage, or undermine, the essential characteristics of the judiciary' (Kirby 2003, p. 44). This requires a critical analysis of the benefits of the use of the mechanism against the 'costs' involved in the operation of

²⁶ In using this division I build upon and expand the division of Pimentel (2009, p. 16-17).

²⁷ This need for institutional legitimacy reflect that oft cited aphorism that 'justice should not only be done, but should manifestly and undoubtedly be seen to be done': *R v Sussex Justices* [1924] 1 KB 256, p. 259 (Lord Hewart). For discussion see Spigelman (2000a, p. 290-292).

²⁸ As Jayawickrama (2002, p. 563) observes, 'the real source of judicial power is the public acceptance of the moral authority and integrity of the judiciary.' Similarly Brody notes that a 'trusted judiciary will be respected and viewed as legitimate. A judiciary that is not trusted may have its legitimacy, authority, and eventually orders questioned by the citizenry or by the other branches of government' (Brody 2008, p. 125).

²⁹ Mohr and Contini (2007, p. 11) argue that accountability can be characterised on the one hand as those systems which instil the appropriate values and interests in an organisation, and, on the other hand, as the mechanisms by which one can assess whether those values and interests are built into that organisations actions. This division creates a fracture between the two conceptions. In contrast, the bifurcation of the two aspects I describe reflects two sides of a common pursuit.

³⁰ Contini and Mohr (2007, p. 30) note that too often in public debates accountability has become an iconic end in itself, with sight lost of its instrumental nature.

³¹ For example, a maladjusted accountability mechanisms may create feedback loops that improperly influence the judge to decide in a 'safe' manner (See Handsley 2001b, p. 182).

accountability mechanisms, including the impact on competing values, time, and financial considerations.³²

4. Judicial performance as a tool of judicial accountability

This instrumentalist conception of judicial accountability profoundly affects the analysis of judicial performance evaluation programs: it requires that we carefully articulate *how* evaluation programs promote the ends of accountability. Performance evaluation must further the pursuit of the internal and external aspects of accountability, and must do so in a way that is of net benefit to the excellent performance of the judicial function.

Judicial performance evaluation programs utilise structured, tightly-regulated and repeatable methodologies to *measure* judicial performance, providing *feedback* on a judge's performance. That feedback can then be utilised to augment other mechanisms of judicial accountability. The provision of feedback does not of itself directly promote the objectives of accountability.³³ It is because of the *consequences* that that may follow from a measurement that such measurement may promote the ends of accountability. It is not the *measurement* of judicial performance that promotes these ends, but the *use* of that information. Internal accountability arises from the act of judicial self-improvement driven by reflection on feedback, and not merely from the measuring of conduct. Public confidence in judicial competence will only be affected where members of the public utilise that information, considering it to form some new and concrete conclusion.

No matter how accurate or extensive a measurement is, it does not have instrumental consequences until it is used in some way. Feedback matters because it can inform subsequent choices and actions which then affect change. Whether the feedback of evaluation is utilised for personnel/management purposes (resulting in promotion, demotions, or even firings) or for developmental purposes (aiding self-improvement) (Felter 2008, p. 158), it is this subsequent use that ultimately matters. Judicial performance evaluation does not operate as a direct mechanism of judicial accountability, as it does not directly affect either aspect of accountability. Instead, performance evaluation acts as a *tool* to provide information to aid the operation of other accountability mechanisms. Judicial performance evaluation indirectly promotes the purposes of judicial accountability by aiding the operation of related primary mechanisms of accountability. The tool of judicial performance evaluation can operate to further the efficacy of a number of accountability mechanisms, including: (1) aiding judicial promotion in professional judiciary; (2) informing judicial retention elections; and (3) aiding professional development and self-improvement.

Ultimately, though, the accountability gain will always be constrained by the limitations inherent in the primary mechanism itself. No matter how well designed or implemented a judicial performance evaluation program is, its role in promoting

³² Indeed care must be taken to ensure that overactive accountability mechanisms do not inadvertently undermine the objectives they should be promoting by too regularly demonstrating that an institution is 'accountable'. As O'Neil argued in her Rieth Lectures, public institutions, like plants that 'don't flourish when we pull them up too often to check how their roots are growing'; they cannot thrive 'if we constantly uproot them to demonstrate that everything is transparent and trustworthy' (O'Neil 2002, p. 6).

³³ While I acknowledge the *possibility* that the mere of measuring performance may directly as a means of accountability, I argue that in most cases there is an anterior process of reflection or critique that ultimately operates to further the ends of accountability. There may appear to be some gain to both the internal and external aspects of accountability merely by measuring judicial performance, and making it be publically known that such measurement is occurring. Programs of judicial performance evaluation may appear to directly promote a public appreciation of the integrity and competence of the judiciary, by showing a judicial openness to critique and demonstrating to the public the competence of the judiciary. Moreover, they appear to promote the internal aspects of actual adherence to integrity and judicial decision-making method by promoting judicial self-reflection. However, it is not the act of measurement itself that is achieving these ends, but the subsequent use of that material.

the ends of judicial accountability will always be dependent upon the use to which it is put. Performance evaluation can make mechanisms of accountability more effective, but it cannot ever rise above the limitations of that mechanism. An instrumentalist conception of judicial accountability demands that judicial performance evaluation be viewed as a derivative tool, operating for accountability purposes but through other mechanisms of accountability.

The derivative instrumental nature of judicial performance evaluation imposes a second profound limitation upon the operation of evaluation programs: the values measured must be rationally connected to the excellent performance of the judicial function. Judges must be evaluated by reference to criteria linked to that such performance. If independence, integrity, and competence are the 'hallmarks' of an excellent judiciary then they must be 'the principles for which a judiciary should be held accountable' (O'Connor 2008, p. 4).

It may be that these values are not capable of direct measurement or indeed of measurement at all. Certainly, Spigelman (2006, p. 72) argues that the core judicial values such as justice, accessibility, openness, fairness, impartiality, legitimacy, participation, honesty, and rationality 'are not capable of measurement, not even by proxy indicators.' He argues that such values cannot be measured, only judged (Spigelman 2002, p. 25). While this conclusion may be debatable, Spigelman highlights a real concern. It does not follow however, that judicial performance evaluation cannot act as a tool to help motivate judges to strive for these values by providing feedback. These arguments do, though, highlight two points: Firstly, evaluation programs must not only measure what is easily measurable, but must create space to 'assess what is important' (Mohr *et al.* 1996, p. 158);³⁴ Secondly, the foundation of any strong judicial performance evaluation program will ultimately rest upon on the quality of the metrics used (Elek *et al.* 2012, p. 72), and the connection of those metrics to the underlying objectives of judicial accountability. Measurement is no neutral thing. Poorly designed measurement criteria can operate to introduce distortive influences into judicial systems (See Spigelman 2002, p. 25-26; Mohr and Contini 2007, p. 35).

The instrumentalist conception of judicial accountability introduces profound restrictions to the operation of judicial performance evaluation programs. However, within those limits it creates a genuine space for such programs to flourish.

5. Examining the implications of the conception

This conceptual framework for the analysis of judicial performance evaluation programs invites us to re-examine existing programs to reflect upon how they respond to these imperatives and limitations. In the previous section I identified three ways in which performance evaluation can promote judicial accountability, namely:

- 1) Aiding judicial promotion in professional judiciaries;
- 2) Informing judicial retention elections; and
- 3) Aiding professional development and self-improvement.

In this section I will examine these different uses by reference to three case-studies presented at, or related to, the Onāti workshop from which this volume emerges. Reflecting on these diverse programs helps demonstrate the benefit of thinking seriously about the nature and scope of judicial performance evaluation.

³⁴ This issue of 'measurement bias' has been described by Drummond in the following terms: 'if something is difficult to measure, it is often treated as unimportant; if it is impossible to measure it is often treated as if it did not exist' (Drummond 2001b, p. 377).

5.1. Performance evaluation and judicial promotions

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.

In a career judiciary, this generally means that appointment to the higher judiciary becomes a matter of promotion from the lower judiciary as it is rational to assume that a candidate who has an excellent judicial track record will make a good higher level judge: past performance is a good indicia of future performance in the new role.³⁵ Moreover, a system of merit-based promotion not only helps to ensure the quality of the higher judiciary, but critically also serves as an effective mechanism of accountability for judges of *all* courts: if judges are promoted on their record of judicial excellence (performance of the judicial function) and judges have a desire to be promoted, then judges will be more motivated to attain excellence. The rewarding of excellence promotes the attainment of excellence. In addition, merit-based promotion mechanisms enhance the perception of judicial excellence, giving the public good reason to have confidence in the ability of the judiciary.

Of course, to relate promotion to the competence of the judge unavoidably requires processes to 'evaluate' the performance of the judge. In Europe judicial performance evaluation has traditionally been seen exclusively as a tool of 'staff management' (Riedel 2014, p. 977), directed to the selection and promotion of individual judges (Contini *et al.* 2014, p. 1102). In his article in this volume, Johannes Riedel, the President of the Cologne Court of Appeal, outlines with great detail and valuable insight the process by which judges are evaluated in this way in Germany. Riedel notes that higher judicial appointments are largely based on the results of performance evaluations (Riedel 2014, p. 983).³⁶ The systematic assessment provided by the programs of judicial performance evaluation provides the crucial tool that enables the merit-based promotion of judges.

5.1.1. Performance evaluation and the problem of proxies

Of course, if evaluation is to aid merit-based promotion in furthering the objectives of accountability, that evaluation must assess *judicial* excellence. If the evaluation criteria are unrelated to the core judicial function then the nexus between evaluation and the objectives of the judicial accountability break down. If a judge is purely evaluated, for example, on the number of cases disposed of, the judge becomes motivated merely to dispose of more cases (See Spigelman 2000b, p. 381, 2006, p. 75). To further the ends of accountability, the focus of assessment criteria must be on underlying values of judicial excellence.³⁷ To promote both the internal and external aspects of accountability, the judge must be assessed on being a good judge.

Of course, it is much easier to list the characteristics of a 'good' judge (Riedel 2014, p. 977-978) than to use systematic evaluation to 'objectively' assess them (See Spigelman 2002, p. 25). To some extent, evaluating judicial performance means

³⁵ Note that this assumption has its limits, and depends upon a sufficient degree of similarity between the roles. It may be, for example, that a given judge is an excellent trial judge, but is poorly suited to the abstractions involved in upper appellate level judicial decision-making. Nevertheless, as a general principle it seems unobjectionable that excellence a lower level is a useful indicator of competence at appellate level.

³⁶ Critically, Riedel concludes that as a criterion of promotion it is 'quite clear that final marks reached in the evaluations play a decisive role in the decision' (Riedel 2014, p. 984).

³⁷ Evaluation does not further the ends of accountability when sight is lost of the goals for which evaluation should be utilised, where 'attention shifts to the minutiae of the data and away from the purpose for which it was required in the first place' (Contini and Mohr 2007, p. 35).

judging judicial decisions. As Riedel (2014, p. 977, 982) notes, this presents the 'unique [difficulty] in evaluating individual judges':

'It is quite clear that judicial independence as guaranteed by the constitution forbids any kind of evaluation which weighs, marks and values the correctness and quality of judicial decisions.'

As a result, German evaluations have been restricted to general observation, assessing a judge for 'personal or professional conduct but never for the way he [sic] has applied the law in his [his] decisions' (Riedel 2014, p. 982). German judicial performance evaluation does not assess the performance of the essential judicial task, instead relying upon a series of proxies to provide a substitute assessment (Riedel 2014, p. 982). Arguably, the conclusion that judicial independence prohibits this form of external evaluation of judicial decisions rests upon a debatable conception of judicial independence that misconstrues the functional nature of independence (See Seibert-Fohr 2010). Nevertheless, putting to one side the normative necessity of this restriction, it is certainly the case that the German judiciary sees this as an unavoidable and meaningful restriction.

The risks inherent in this approach are, however, significant. Most troublingly, the use of performance proxies risks sundering the connection between performance evaluation and the accountability objectives of merit-based promotion. To further those objectives, promotion mechanisms must motivate the judge to perform the core judicial function with excellence and integrity. The use of proxies that exclude consideration of the judge's substantive decision-making pose a three-fold risk. Firstly, the link between judicial merit (in the sense of the excellent performance of the judicial function) and judicial promotion ceases to be the defining feature of evaluation. 'Merit' becomes assessed by reference to the proxy values. The 'best' judge for promotion purposes is the judge that 'best' reflects those proxies in his or her judicial conduct. To the extent that 'merit'-promotion affects judicial behaviour, this method of evaluation risks motivating judges to pursue these proxies rather than the direct pursuit of excellence, thereby undermining the very mechanism of accountability it exists to support. While the chosen proxies will ideally approximate or indicate good performance, such a relationship is contingent and vulnerable. The good judge may decide cases without undue delay, but that does not mean that judicial excellence can be equated with a high case-disposal rate. Such proxies risk judges altering their conduct to increase chances for promotion. Any argument that judges of integrity would resist such a temptation only illustrates the point: the use of proxies creates a temptation that can distort judicial performance. Secondly, and related to this point, there is the risk that by substituting an external motivation value (promotion) for the underlying integrity and internalised self-motivation of judge that such mechanisms can undermine the integrity of the judge. The pursuit of promotion risks being seen as 'greedy' and 'self-serving', thereby undermining the personal integrity and duty-driven attainment of excellence that are the hallmarks of true professionalism. Finally, the use of proxies undermines the external confidence benefits of merit-based promotion. Where merit is assessed by direct reference to the excellent performance of the judicial function, merit-based promotion gives the public good reason to be confident in the ability, competence and integrity of the judiciary. The use of proxies greatly weakens this connection, for the same reason that it undermines the actual attainment of excellence.

For these reasons, the use of proxies risks distorting and fracturing the connection between merit-based promotion and the underlying purposes of judicial accountability. By indirectly measuring performance in this way, performance evaluation risks undermining the mechanisms it is designed to serve. One response to this threat is to abandon merit-based promotion. Practices of promotion on the basis of seniority have chosen this route, though it is beset with its own difficulties. Alternatively, we may choose to abandon the tool of performance evaluation on the basis that the distortive risks of gamesmanship outweigh any accountability

gains.³⁸ Alternatively, it may be that substantive judicial performance constitutes an informal factor that influences merit-based promotion. There may be an element of cognitive dissonance, whereby judicial performance influences evaluation despite explicit commitments to the contrary. Fourthly, we may recognise the real potential of these threats to undermine the objectives of accountability, yet nevertheless decide that the benefits of structured evaluation outweigh these threats. This approach represents an uneasy compromise, and is only sustainable if honest, informed and regular reflection on the entirety of the program is undertaken. Finally, and perhaps preferably, we can lessen the role of evaluation feedback in merit-based promotion to allow space for genuine reflection and appraisal of the quality and competence of the judges' judicial performances. Performance evaluation would inform promotion decisions, but would not be determinative in the way Riedel describes. This approach directly challenges the proposition that any reflection on substantive performance undermines judicial independence. However, if the quality and competence of a judgment can be assessed independently of its substantive outcome (substantive correctness), then such evaluation would not interfere with judicial independence.³⁹ This approach is arguably more intellectually honest than the other options above, though all of these approaches have their difficulties.

Judicial performance evaluation can be a useful aid the operation of merit-based promotion mechanisms of judicial accountability, informing promotion choices and motivating judges to attain excellence in their judicial role. However, such programs can also distort the operation of promotion mechanisms, undermining accountability and the attainment of excellence. This risk requires reassessment of both the theory and practice behind the use of evaluation programs.

5.2. Performance evaluation informing judicial retention elections

Secondly, judicial performance evaluation programs have, since their genesis, been used to enhance the operation of judicial retention elections. The practice of subjecting judges to elections for appointment or retention is an almost exclusively US practice. Some ninety per cent of America's state judges must stand for election, either to gain or retain office (Brandenburg 2009, p. 372). Judicial performance evaluation is utilised in those states which augment merit selection of judges with subsequent retention elections.⁴⁰ The rationale underlying retention elections as a mechanism of judicial accountability is that as judges' re-election will depend upon judicial performance (See Tarr 2009, p. 610), then judges will be motivated to perform the judicial role with excellence. If retention elections actually made judges' tenure dependent upon excellent judicial performance, then they would provide a powerful motivation for judges to act with integrity and excellence and would promote the external confidence objectives of judicial accountability. Retention elections can operate as effective tools of judicial accountability.

5.2.1. Performance evaluation and the provision of information to voters

However, if judges are to be assessed by voters on their judicial performance, it is necessary for the electorate to have access to relevant information about that performance. Judicial performance evaluation programs can fill the 'information vacuum' created by retention elections to lend credibility and meaning to those elections (Pelander 1998, p. 647). These programs strive to comprehensively

³⁸ This is the response of Colbran when he states: 'There is no viable link between judicial performance evaluation and judicial promotion... The potential for increased productivity is more than outweighed by the intrusion on judicial independence, and the risk of gamesmanship distorting the function of the judiciary in dispensing justice' (Colbran 2001, p. 357).

³⁹ I recognise that this approach depends upon a particular conception of the nature of judicial independence, and upon a particular analytic jurisprudential approach to the concept of law and of judicial decision-making that are not without controversy. It is beyond the scope of this paper to engage in such weighty debates.

⁴⁰ As Pelander (1998, p. 647) notes, retention elections are 'an integral aspect of merit selection'.

evaluate each judge's performance, and distribute that information to the voting public (Berch and Bass 2014, p. 932-933). By informing the public about the performance of sitting judges (Brody 2000, p. 341), judicial performance evaluation is seen to add a 'powerful layer' of to the operation of retention elections as a mechanism of judicial accountability (Berch and Bass 2014, p. 932-933).⁴¹ The feedback provided by such programs becomes the 'missing ingredient' necessary for 'judicial elections to appropriately and effectively facilitate judicial accountability' (Brody 2008, p. 129). Judicial performance evaluation provides the 'objective, survey-based information on the performance of judges' (Anderson 2001, p. 1375) that is necessary for the accountability mechanism of retention elections to serve its purpose. The assumption is that a judge with good evaluations will have a far better chance of being retained.⁴² Official judicial performance evaluation programs, operating for the express purpose of enabling voters to cast intelligent, meaningful votes in retention elections, have now become reasonably common.⁴³ Such programs generally employ five evaluative criteria: knowledge of the law; integrity and freedom from bias; clarity of communication; judicial demeanour; and administrative capacity (Kourlis and Singer 2009, p. 657). These programs are generally centered on responses to standardized, scaled surveys provided by individuals who have had direct dealings with the judge.⁴⁴ Such programs are seen as an effective way of reconciling the demands of independence and accountability in the US context (See White 2002, 2009, Kourlis and Singer 2007, Brody 2008).

In her article in this volume Chief Justice Rebecca White Berch of the Arizona Supreme Court provides an overview of the scope and operation of the Arizona system of 'judicial performance review' (Berch and Bass 2014, p. 930; See also Pelander 1998, p. 672). After outlining the history and processes of the program (Berch and Bass 2014, p. 931-936), she usefully examines both its successes and weakness. Interestingly, she argues that possibly the greatest success of the program is its impact upon self-evaluation and improvement aspects (Berch and Bass 2014, p. 936-937), which was not the original rationale behind the program. Secondly, Berch sees as a major success the evidence that the Commission's information regarding Arizona's judicial performance evaluation program is reaching voters.⁴⁵ The performance evaluation program does seem to provide more information to voters, and voters seem to be willing to look for that information.

5.2.2. Criticism of evaluation programs

Berch notes, however, issues of concern in the operation of Arizona's judicial performance evaluation program. These concerns reflect common criticisms of corresponding evaluation programs. These criticisms fall into three broad categories: (1) concern over the 'accuracy' of any evaluation; (2) concern over the use of that information by voters; and (3) concern over the potential of the evaluation program to distort judicial behaviour. I will briefly address each of these concerns.

1. Inaccurate assessments

Firstly, there is a concern that the assessments provided by the evaluation programs do not accurately reflect the judicial performance and competence of the judges. One concern is that such assessments systemically 'over-rate' judicial

⁴¹ As Brody notes, while 'judicial elections provide the vehicle to provide judicial accountability to the public and to provide true accountability that will increase independence, voters need to be given sufficient information with which to make electoral decisions' (Brody 2008, p. 126).

⁴² As Tinkham notes, a judge 'with good evaluations is unlikely to be opposed in the next election, while one with poor evaluations is more likely to be opposed' (Tinkham 2011, p. 1649).

⁴³ As of 2008, eight states provide their citizens with information obtained from judicial performance evaluation programs for retention election purposes (Brody 2008, p. 118).

⁴⁴ For discussion of process generally, see Brody (2008, p. 119).

⁴⁵ For example, the 'hits' on the Commission's website, where the information on performance evaluation is held, went from a long term average of 100/day to an average 16500 hits a day in weeks leading up to the election (Berch and Bass 2014, p. 938).

performance.⁴⁶ Whether or not this concern is misplaced (See Berch and Bass 2014, p. 939), it does highlight the inevitable subjective aspect to all evaluation. Contrary to the commonly perpetuated 'myth', evaluation programs unavoidably contain significant discretionary elements that can distort the accuracy of any findings.⁴⁷ This issue of accuracy is evident in the second area of concern, namely the failure of attorneys to provide full and honest evaluation in judicial surveys (Berch and Bass 2014, p. 940). The concern is that attorneys will 'under-rate' judges they 'deem bad for business' (Berch and Bass 2014, p. 940).⁴⁸ The potential of practitioner surveys to provide, as a result of respondent self-interest, a distorted assessment of judicial performance was precisely the criticism levelled against traditional 'Bar Polls'. This threat has not been eliminated in the context of state regulated performance evaluation programs. Thirdly, there is concern that evaluation programs may be systematically biased against particular groups of judges. Empirical research does suggest that women and members of minority groups receive more negative evaluations than their white male counterparts (Elek *et al.* 2012, Gill 2012). Such findings of systemic bias not only undercut the accuracy of evaluation findings, but highlights that the foundation of any strong performance evaluation program rests most heavily on the quality of the metrics used (Elek *et al.* 2012, p. 66). These metrics must be intimately connected with the characteristics desired in the judge. That connection is not always present. Common to all these concerns is the recognition that performance evaluation is only useful for promoting the accountability objectives of retention elections so long as it provides an accurate assessment of judicial performance. Such programs can only promote accountability where they provide information of a sufficient quality and accuracy. This is not always the case.

2. Limited use of information

The second group of criticisms concerns the use, by voters, of the information provided through evaluation programs. Retention elections will only be effective mechanisms of accountability where engaged and informed voters are willing to meaningfully reflect upon the performance of judges and vote according to judicial competence. There is evidence that this is not the case, irrespective of the quality of performance information provided. Firstly, there is strong evidence that there remains significant 'under-vote' with respect to retention elections. Berch and Bass note an average 'under-vote' of 42.9% – voters who submitted a ballot but did not cast a vote for a particular judge (Berch and Bass 2013, p. 940). This issue appears to be a recurrent theme in judicial retention elections (Pelander 1998, p. 711). Perhaps more troublingly, there is evidence that such elections remain marred by voter's lack of knowledge about judicial candidates (Pelander 1998, p. 711). Evidence of uninformed voters abounds (Berch and Bass 2013, p. 940). For example, many voters continue to treat all judges on the ballot the same, voting either for or against all judges (Berch and Bass 2013, p. 940). This pattern is reflected in other jurisdictions.⁴⁹ Moreover, it seems that even the voters who do differentiate between judges largely ignore evaluation recommendations (Berch and Bass 2013, p. 941). Griffith (1995, p. 62) suggests that there is little evidence to suggest a favourable report will influence these voters (see also Pelander 1998, p. 709). Equally, there is strong evidence that voters will even routinely ignore

⁴⁶ Berch and Bass note, for example, the concern that the Commission is reluctant to vote that a judge 'Does Not Meet' the performance standards (Berch and Bass 2014, p. 939).

⁴⁷ Spigelman notes that opinion surveys about quality are based on perceptions, and that such surveys 'are notoriously unreliable, particularly on matters about which the interviewees have limited understanding' (Spigelman 2006, p. 78).

⁴⁸ Evidence of this practice is provided in the example given by Brody of the 'hard-nosed' judge who was deliberately appointed to bring order to the criminal docket only to be given a low rating for being hard-nosed (Brody 2000, p. 344-346).

⁴⁹ Griffin (1995, p. 62), for example, notes that approximately 30% of the electorate 'routinely votes 'no' in judicial retention elections, no matter who the judge happens to be', nor how glaring or effusive the Commission's evaluation of the judge is. See also Pelander (1998, p. 709).

negative findings.⁵⁰ No matter how accurate the information provided by judicial evaluation programs is, unless that information is being actively used by voters in their electoral choices, those programs will be of little utility. Evidence of voting patterns illustrates a real concern as to the use of evaluation reports, giving good reasons to doubt the efficacy of even well-implemented programs. Where performance evaluation becomes disconnected from the overarching accountability mechanism of retention elections its use becomes unjustifiable.

3. Potential for undue influence

Finally, there remains the potential for the operation of performance evaluation programs to improperly influence judicial performance. While this is not a concern Berch notes in respect of Arizona, other commentators have certainly raised this spectre.⁵¹ The concern is that judges may alter their behaviour to unduly gain a favourable report, with the consequence that the evaluation programs may actively distort judicial performance thereby undermining the internal objectives of accountability. Moreover, while evaluation programs can provide the public with a sense of oversight and control over the judiciary, if the program is viewed as corrupt, slanted, or unfair, its existence may undermine public trust in the judiciary (Brody 2000, p. 335). Any perception of potential undue influence will diminish public confidence in the judiciary, undermining the external objective of judicial accountability. Judicial performance evaluation is a tool to aid the accountability mechanisms of retention elections. There is inevitably a cost in its use that may ultimately outweigh any benefit gained with respect to the overarching mechanism.⁵² If the cost of using performance evaluation programs, in terms of actual undue influence or the perception of such, outweighs any benefit to the operation of the accountability mechanism then its use cannot be justified.

It can be seen that there is real concern over the efficacy, accuracy and justifiability of judicial performance evaluation in the context of judicial retention election. Ultimately though, these concerns may not illustrate problems with the tool of performance evaluation, but with the primary mechanism of retention elections.

5.2.3. Critique of the primary mechanism

There is growing concern that judicial retention elections may no longer be serving the accountability purpose for which they were originally developed (See Geyh 2003, Tarr 2009): a form of 'ballot box accountability' whereby judges who do not perform their judicial role with excellence, competence, and integrity will not be retained. If retention elections did in fact operate in this way then they could provide a powerful means of motivating judges to pursue such qualities.

However, there is evidence to suggest that judicial elections are increasingly, about politics and the attempt to influence judicial decision-making rather than judicial merit. This is notoriously illustrated by the removal of Penny White from the Tennessee Supreme Court following a concerted attack against her as a result of her participation in a majority decision on a death sentence appeal (See Anderson 2001, p. 1378-1379, Brody 2008, p. 123-124).⁵³ Such instances constitute a direct attack on judicial independence (Brody 2008, p. 124): arguably judicial elections

⁵⁰ Paynter and Kearnes (2010, p. 943) note that in the six states they surveyed, there were 28 instances of judges receiving 'Do Not Retain' recommendations. Of these judges, 13 were not retained, while 15 were retained.

⁵¹ Griffin, for example, argues that so long as a judge's right to remain a judge is dependent (or even only influenced) by the will of a committee of 10 to 15 other people, then 'sooner or later judicial behaviour will be affected' (Griffin 1995, p. 7).

⁵² In this context, it is worth also considering the financial costs involved in operating formal judicial performance evaluation programs. Berch notes that Arizona's program costs approximately \$269,300 annually, though notes that this figure provides only an estimation of time spent by court staff and does not factor in the countless hours spent by volunteers who serve on the Commission or on the self-improvement teams (Berch and Bass 2014, p. 13). See also Pelander (1998, p. 706-707).

⁵³ Other notorious examples include the successful campaign against the retention of Chief Justice Rose Bird (together with justices Cruz Reynoso and Joseph Grodin) in California in 1986.

have become have the 'primary vehicle' for retaliation against judges on account of their decisions (Geyh 2003, p. 49). Where votes are influenced by the voters opinion of particular judicial decisions, there a potential for judicial elections to 'pervert the decision-making process it proposes to improve' (Spigelman 2006, p. 79). Judges may become incentivised to improperly alter their judicial decisions in order to promote their chances of being re-elected, potentially abandoning what they believe to be correct in favour of what they believe is popular. This is the very antithesis of what judicial accountability mechanisms are supposed to achieve.⁵⁴

It is improper for any person or interest group to attempt to unduly influence substantive judicial decision-making through retention elections.⁵⁵ Such interference fundamentally challenges judicial independence. The judicial values of fierce independence and adherence to the law may sit uncomfortably with populist forms of accountability, yet this tension between majoritarianism and legalism ought not be quickly dismissed.⁵⁶ This tension lies at the heart of any attempt to hold judges to account through electoral means. If we truly want 'judges' who are 'politically accountable to the political clamour of the moment' then the decision-makers we aspire to are not performing a judicial function (See Feltner 2008, p. 179).⁵⁷ If, instead, we want judges who aspire to the excellent performance of the judicial function, then populist-based accountability for judges should be feared and avoided (Kourlis and Singer 2008, p. 8). The purpose of judicial accountability remains the promotion of judicial quality and excellence. Though it may be unfashionable to say so, quality is, as Spigelman notes, 'by its nature, not susceptible to democratic assessment' (Spigelman 2006, p. 79).⁵⁸ There are good reasons to suggest that retention elections no longer serve the purposes for which they were created (Tarr 2009, p. 632).

I do not wish to fall into the trap of the outsider criticising a quintessentially American practice from afar. Instead, I raise these concerns to illustrate a fundamental limitation on the use of judicial performance evaluation programs. Where such programs exist to aid the performance of primary mechanisms of accountability, then no matter how well designed and implemented those programs, their utility will always remain restricted by the inherent limitations of that primary mechanism. If the primary mechanism is poorly suited to the promotion of the objectives of judicial accountability, then not even a perfect evaluation program can be justified on accountability grounds. The evidence of voting patterns suggests that retention elections are an ineffective means of motivating judges to perform their judicial role with excellence. Likewise that evidence suggests that too often the judicial competence of the judge is irrelevant to voters. In this context, retention elections provide minimal incentive to the judge to pursue excellence. Given the other distortive potential involved in this process, there may be good reason to doubt the utility of such elections. As a result, it would follow that judicial evaluation programs could not be justified by reference to their influence on retention election voting. The stream cannot rise above its source, and the utility of the tool will always remain limited by the ability of the primary mechanism to promote the objectives of judicial accountability.

⁵⁴ These concerns are compounded by the rapid escalation of the money that is being poured into judicial election campaigns (See Brandenburg 2009, p. 373). This money creates new avenues by which judges can become improperly influenced by, or beholden to, external interests.

⁵⁵ As Brody notes, the basis upon which voters hold judges accountable 'must not infringe upon their ability to decide matters based on the law' (Brody 2008, p. 126).

⁵⁶ Brody rightly identifies this tension: 'Making anti-majoritarian decisions, developing innovative practices and procedures, and standing by one's beliefs – all qualities we want from our judges – can make a judge liable to electoral challenges based on unfair characterization, innuendo, and attack by special interests' (Brody 2004, p. 168).

⁵⁷ As Feltner notes, if we want such 'politically accountable' decision-makers it may be better to have a legislative body, which represents constituencies of the public, vote on the resolution of specific controversies between litigants' (Feltner 2008, p. 179).

⁵⁸ Spigelman notes that 'quality is hard to judge. It requires knowledge and experience' (Spigelman 2006, p. 79).

5.3. Performance evaluation and judicial professional development

Finally, judicial performance evaluation can operate as an effective aid for structured judicial professional development programs. As the instrumental end of judicial accountability is not to 'hold to account' judges, but to promote judicial integrity and the excellent performance of the judicial function, professional development programs can be effective mechanisms of judicial accountability. The opportunity for meaningful self-reflection and supported improvement provided by such processes can promote the objectives of accountability. The internal objectives of accountability, of promoting judicial integrity and the pursuit of judicial excellence, are effectively served by programs directly designed to educate judges and encourage self-improvement. Likewise, knowledge of such programs gives the public good reason to have confidence in the ongoing competence and quality of the judiciary.

Of course, there is a legitimate expectation that judges, as professionals, will possess a well-developed capacity for independent learning (Goldring 2003, p. 216). This internal drive for improved judicial performance is a critical component of judicial integrity, and complements the honesty, good faith, and diligence judges are expected to bring to the bench (Aynes 1981, p. 255). However, there remains a significant skill-based aspect of judicial performance, the improvement of which can have a profound impact on the quality of justice and the overall efficacy of the judiciary (Aynes 1981, p. 255). Well-structured and organised development programs aid the enhancement of those skills by providing an opportunity for self-reflection and improvement (Warren 2011, p. 5). Such programs help judges internalise the standards of professional conduct expected of the judge, and invite judges to reflect whether they are fulfilling their professional obligations (Goldring 2003, p. 219).

5.3.1. The role of performance evaluation in informing professional development

Judicial professional development is made difficult by the fact that in the normal course of their role judges receive little informed feedback on their judicial performance. The very nature of their work means that judges, particularly in the common law world, often work alone as individual decision-makers. They seldom have extensive external stimuli, and are 'virtually unique' in the paucity of feedback they receive on their professional performance (Goldring 2003, p. 217). While some review of substantive evaluation of decision-making is provided through appellate supervision, that system is 'singularly ill fitted to provide constructive feedback to judges' (Goldring 2003, p. 217). Existing conventions mean that judges' peers do not observe them operating in a professional capacity, and are therefore in no position to provide feedback on their performance (Goldring 2003, p. 217). As Chief Justice Marilyn Warren of the Victoria Supreme Courts notes (Warren 2011, p. 5):

'...the unintended consequence of the environment in which we work is that we receive little or no formal feedback about how each of us perform the role. As judges, it is a very lonely and isolated life that we lead. We sit on the bench day in day out, and from the very beginning everyone is very polite in court. But no-one ever tells us how we are going, how we are performing. ... We have, as judges, no feedback.'

Judicial performance evaluation programs can respond to this void by providing judges with reliable feedback that can materially assist in that process of professional development (Griffin 1995, p. 6): areas for improvement are identified, and areas of excellence recognised (Aynes 1981, p. 255). A well-implemented judicial performance evaluation program can address the problems of isolation or a lack of self-reflection (Haines 2010, p. 926) that can limit the ability of judges to hone their judicial skills. The feedback such programs provides can help judges make informed choices as a basis for self-improvement (Colbran 2001, p. 356), and allow the judge to become aware of perceived weakness without fear of public

embarrassment or retaliation (Haines 2010, p. 926). This feedback aids the internal drive to self-improvement, assisting the judge to achieve a higher degree of judicial competence. In doing so, performance evaluation strengthens the efficacy of the primary mechanism of accountability of the professional development programs, helping to improve the judiciary by fostering judicial self-development.⁵⁹

This potential for judicial performance evaluation programs to provide feedback for judicial self-development, thereby promoting the excellent performance of the judicial function, is well illustrated by the 'Court Craft' program implemented in Victoria. That program, described by Warren, complements a '360-degree feedback' model with surveys and interactive training to provide detailed feedback to judges on their judicial performance and demeanour (See Warren 2011). The stated intention of the program is to provide an opportunity for judges to gain insight into their judicial performance in a way that enhances on-going opportunities for professional development (Warren 2011, p. 5). The program has been, Warren (2011, p. 6) argues, both a 'delicate and very successful process'. That term 'delicate' highlights that the success of the program is, as Warren (2011, p. 7) acknowledges, 'undeniably linked to the fact that participation is on a voluntary basis and that it is a very affirming process.' Such programs require judicial 'buy-in': no matter how 'accurate' the feedback provided by an evaluation program, that program can only succeed in aiding professional self-development where the judges willingly embrace the program and utilise the information provided by it.

5.3.2. Limitations and criticisms

This critical need for 'buy-in' by the judiciary is a limitation inherent in any program of judicial performance evaluation for professional development purposes (Haines 2010, p. 926). The degree to which a judge will benefit from such a program will depend greatly, if not entirely, on the judges' attitude toward the process (Berch and Bass 2014, p. 937). Judges are human, and like all people dislike being judged and criticized (Paynter and Kearney 2010, p. 927). Not only can this lead to judges closing their mind to a useful source of feedback, but can develop into resistance to participation in such a program, which if wide spread, will seriously undermine the program.⁶⁰ Other criticisms of the use of judicial performance evaluation for professional development purposes include issues of cost, confidentiality and undue influence. Firstly, such programs can be difficult and expensive to design and implement, consuming significant amounts of employee and organizational time (Paynter and Kearney 2010, p. 929). Secondly, there may be problems ensuring and maintaining the confidentiality and anonymity of respondents (Paynter and Kearney 2010, p. 929); without faith in the confidentiality of the process, any feedback is less likely to be honest and accurate. Thirdly, there is concern that any formalised judicial education program may potentially exert undue influence on judges through structural and systemic biases in teaching methodologies. However, in contrast to situations where performance evaluation is linked to salary or promotion, the potential for gamesmanship appears minimal in a professional development context (Colbran 2001, p. 356).

By understanding performance evaluation as a tool to aid the operation of mechanisms of judicial accountability, we are forced to think more clearly about the costs and benefits of the tool. Where evaluation is utilised for professional development purposes, those benefits will generally outweigh usage 'costs'. Structured evaluation can provide meaningful and valuable feedback to judges

⁵⁹ It should be recalled that this self-improvement focus was the purpose for which the ABA originally advocated for judicial performance evaluation and is, according to Berch and Bass, the most successful aspect of Arizona's program (Berch and Bass 2014, p. 936-937); see also Pelander (1998, p. 647, 708).

⁶⁰ I note that for some judges of high integrity and competence, such programs may not accord with their own methods of development, and accordingly they will gain little from it. Not all judges will benefit from such programs, but there is a significant difference between an informed and reflective desire not to participate, and a defensive rejection of the program.

about their judicial performance,⁶¹ while the risks inherent in utilising judicial performance evaluation for professional development purposes appear significantly less than for other uses of the tool. The connection between the feedback role of such evaluation and the accountability objective of the professional development mechanisms is much more direct than in other uses of performance evaluation.

6. Conclusions

Over the last 40 years, the use of 'judicial performance evaluation' has grown from a radical idea to a familiar concept. It is now possible for judges and scholars of judicial studies from around the world to gather to meaningfully discuss this common phenomenon. Nevertheless, there remain distinct approaches to the methods of judicial performance evaluation and, perhaps more importantly, the purposes for which those programs are used.

In this article I have argued that there is a common weave that links together these distinct practices in both the form and purpose of judicial performance evaluation. By its form 'judicial performance evaluation' refers to the use of structured, tightly-regulated and repeatable methods to *measure* judicial performance to provide *feedback* on a judge's performance in a way that does not depend upon mere opinion of a few (often partial) individuals. The common aim of that measurement is to promote the objectives of judicial accountability. In this article I argue that to understand the modern practice of judicial performance evaluation, and the limitations of evaluation programs, it is necessary to understand the relationship between such programs and the principles of judicial accountability.

6.1. A Conceptual framework for analysing judicial performance evaluation programs

A conceptual analysis of the purposes and limitations of judicial performance evaluation has largely been lacking in the literature. For the most part, academics have been content to examine and critique particular programs in a specific jurisdiction (compare, for example, Spigelman 2002, 2006, and Drummond 2001a, 2001b, with Griffin 1995, Colbran 2002b, and White 2002). Even where there is quality comparative analysis (See Mohr and Contini 2007, Contini and Mohr 2007), it largely focuses on particular uses of performance evaluation. This focus on discrete mechanisms and on particular jurisdictions means that, by and large, systematic assessments of judicial performance evaluation programs, and their impact on judicial accountability and independence, have been lacking (Brody 2008, p. 120). At a deeper level, the ability to analyse the impact of judicial performance evaluation on judicial accountability has been severely hampered by the paucity of quality theory with respect to judicial accountability itself.

I begin, therefore, by examining the concept of judicial accountability. I argue that judicial accountability is an instrumental concept that operates for the limited purpose of promoting the excellent performance of the judicial function. This involves both internal aspects (promoting actual judicial integrity, excellence and quality), and external aspects (promoting public confidence in, and a reputation for, such values). Mechanisms of judicial accountability can promote these objectives in a variety of ways but are bound by this instrumental nature; where their operation undermines the excellent performance of the judicial function their use cannot be justified.

⁶¹ While it may put it too strongly to suggest that without judicial performance evaluation judges 'are not in a position to objectively evaluate their own performance and skills' (Colbran 2002b, p. 239, 2003b, p. 68) this feedback clearly has the potential to support more effective judicial professional development. It should be noted that Colbran's assertion makes a number of questionable assumptions, including that judicial performance evaluation is in some way 'objective' and that this is desirable; that evaluation programs are capable of delivering such 'objective' feedback; and thirdly, that only formalised process of JPE are capable of deliver such feedback.

Judicial performance evaluation operates to aid the operation of various primary mechanisms of judicial accountability. By providing detailed feedback on judicial performance, these primary mechanisms can operate more effectively. However, there is a cost in the use of the tool of evaluation, and that cost must be balanced against any gains provided. Moreover, the derivative nature of judicial performance evaluation means that no matter how well designed or implemented a program may be, its utility will always remain limited by the overarching accountability mechanisms it serves.

6.1.1. application of framework

I then go on to illustrate the utility of this approach by critically analysing three of the more significant roles for which judicial performance evaluation is utilised: judicial promotions; retention elections; and professional development. The first issue illustrated how the process of measuring performance can operate as an improper influence on judicial decision-making, distorting the very performance the mechanisms is supposed to promote. That threat was particularly acute in the program described by Riedel (2014) which explicitly avoids any attempt to assess substantive judicial performance. The use of proxies creates the potential for gamesmanship that frustrates the pursuit of excellence in judicial performance at the heart of judicial accountability. The apparent 'objectivity' of performance evaluation may undermine the very mechanism of merit promotion it is supposed to support. The inverse problem arose in the context of judicial retention elections, where the evidence suggests that far too often the voters will disregard or ignore the evaluative feedback provided to them. Irrespective of the quality of the evaluation program, the limitations inherent in the primary mechanism of retention elections will fundamentally restrict the utility and impact of the program. Performance evaluation remains a tool, limited by the overarching mechanism it serves. Finally, the professional development context shows the benefits that can arise where the tool of performance evaluation harmonises with the accountability objectives of the primary mechanism. The feedback for professional development provided by voluntary evaluation programs with a high degree of 'buy-in' can provide substantial benefits for low functional costs. Such evaluation effectively promotes the objectives of judicial accountability with minimal drawbacks.⁶²

I use these examples not to assert the appropriateness or otherwise of a particular use of judicial performance evaluation in a given particular jurisdiction. Rather these examples illustrate the way in which a clear conceptual framework enables such issues to be answered. I argue that by conceptualising judicial performance evaluation as a derivative tool that operates to promote mechanisms of judicial accountability, we are far better placed to understand the strengths and weaknesses, purposes and limitations of such evaluation.⁶³ The examples I give help to illustrate the utility of this approach. There may indeed be other accountability purposes served by performance evaluation, particularly for matters of judicial and court management.⁶⁴ Alternatively, it may be argued that judicial performance evaluation operates directly as a mechanism of accountability, promoting the external objectives of accountability of public confidence merely by

⁶² Indeed, it has been suggested that in a context like Australia, judicial self-improvement is the *only* appropriate use for judicial performance evaluation (Colbran 2001, p. 336). I do not make such bold normative claims in this paper. While there are good reasons for accepting such a claim, it goes beyond what I have demonstrated here.

⁶³ Moreover, this clarity can assist in designing the final form and mechanics of any program of judicial performance evaluation. As Aynes notes where 'the goal of a performance evaluation program is to assist in the professional development and improvement of sitting judges, a totally different methodology may be necessary than where the purpose is for promotion or election' (Aynes 1981, p. 269).

⁶⁴ This can operate at either the level of the allocation of resources to various courts and court offices or in some cases for the remuneration of personnel (Contini and Mohr 2007, p. 37).

the act of measurement and public knowledge of that measurement.⁶⁵ Such alternative uses do not challenge the framework I have developed, but rather invite its further application.

6.2. Final observations on evaluation programs

The purpose of this article is to examine the analysis of judicial performance evaluation programs from a different angle. I do not, here, comment on the appropriateness of particular proxies or mechanisms. Rather, I address the largely unspoken prior question: why are we evaluating at all? This invites us to think critically about judicial performance evaluation programs, focusing on their objectives, purposes and effects rather than simply their mechanics. The modern concept of judicial performance evaluation is, at its core, about the *measurement* of judicial conduct, performance and behaviour. The process of measurement invites several questions, including:

- 1) What is the behaviour or value to be measured?
- 2) Can the behaviour or value be meaningfully measured?
- 3) Can that measured value serve a given purpose?
- 4) What are the costs of measurement? and
- 5) Does the measurement alter that which is measured, distorting the underlying practice?

The framework I have developed helps answer some of these questions: the first three of these issues help us to think about what is measured and why. The final issues look at the implications of that measurement. Underlying this is a question regarding the meaningfulness of measuring judicial performance *at all*. Such measurement aims to replace subjective assessment with more 'objective' evaluative criteria. Judicial decisions are, however, dependent upon acts of *judgement* by the judge, upon unavoidable and genuine exercises of discretion. Subjectivity is arguably a necessary and unavoidable aspect of judicial evaluation. As Spigelman notes, justice and judicial performance may be one to those things that cannot be measured, only judged (Spigelman 2002, p. 25). Objectivity may not only be an illusion in this context,⁶⁶ but attempts to pursue it may sideline opportunities to examine key aspects of, and motivations for, judicial performance.⁶⁷

Understanding the accountability purposes of performance evaluation highlights that the proper focus of *judicial* performance evaluation is on the excellent performance of the judicial function, not merely on those things that are readily measurable. This underlying accountability purpose not only guides and informs any attempt to answer these five questions, but more fundamentally helps to expose the proper purposes and limitations of judicial performance evaluation.

References

ABA, 2007. *ABA Model Code of Judicial Conduct, February 2007* [online]. Chicago: American Bar Association. Available from:

⁶⁵ The mere existence of an evaluation program *may* increase confidence in the courts. As Haines notes, even where 'an evaluation is confidential, if the public is aware of the fact that a judge is subject to this kind of review, they are hard pressed to argue that judges are out of touch or acting on their own without any guidelines' (Haines 2010, p. 926). Moreover, it may strengthen confidence by informing the public of the proper role of courts (Brandenburg 2009, p. 379).

⁶⁶ Spigelman notes that quantitative measurement 'appears to be objective and value free. Qualitative assessment appears to be subjective and value laden. In fact statistics, by reason of their selectivity, contain and conceal important value judgment' (Spigelman 2002, p. 25).

⁶⁷ Of course, recognition of the role of genuine evaluation does not isolate the judiciary from informed critique. Indeed, the need for judgement invites a wide ranging discourse on what we desire in the performance of our judiciary in a way that is unlikely to be present when myths of objectivity are maintained.

http://www.americanbar.org/content/dam/aba/migrated/judiciaethics/ABA_M_CJC_approved.authcheckdam.pdf [Accessed 22 December 2014].

- Anderson, S., 2001. Judicial Retention Evaluation Programs. *Loyola of Los Angeles Law Review* [online], 34 (4), 1375-1390. Available from: <http://digitalcommons.lmu.edu/llr/vol34/iss4/7/> [Accessed 3 December 2014].
- Aynes, R., 1981. Evaluation of Judicial Performance: A Tool for Self-Improvement. *Pepperdine Law Review* [online], 8 (2), 255-312. Available from: <http://digitalcommons.pepperdine.edu/plr/vol8/iss2/1/> [Accessed 3 December 2014].
- Bandes, S., 2006. Judging, Politics, and Accountability: A Reply to Charles Geyh. *Case Western Reserve Law Review*, 26 (4), 947-964.
- Berch, R.W. and Bass, E.N., 2014. Judicial Performance Review in Arizona: A Critical Assessment. *Oñati Socio-legal Series* [online], 4 (5), 927-952. Available from: <http://ssrn.com/abstract=2533868> [Accessed 23 December 2014].
- Brandenburg, B., 2009. Courts Under Pressure: The Growing Threat To Impartial Justice. *Syracuse Law Review*, 59 (3), 371-382.
- Brennan, N., *et al.*, 1986. Judicial Performance Evaluation. *Law and Inequality*, 4 (1), 77-90.
- Brody, D., 2000. Judicial Performance Evaluations by State Governments: Informing the Public While Avoiding the Pitfalls. *The Justice System Journal* [online], 21 (3), 333-347. Available from: http://www.americanbar.org/content/dam/aba/migrated/jd/lawyersconf/performanceresource/article/Brody_Justice_System.authcheckdam.pdf [Accessed 3 December 2014].
- Brody, D., 2004. The Relationship Between Judicial Performance Evaluations and Judicial Elections. *Judicature* [online], 87 (4), 168-177. Available from: http://www.americanbar.org/content/dam/aba/migrated/jd/lawyersconf/performanceresource/article/Brody_Judicature_87_168.authcheckdam.pdf [Accessed 3 December 2014].
- Brody, D., 2008. The Use of Judicial Performance Evaluation to Enhance Judicial Accountability. *Denver University Law Review* [online], 86 (1), 115-56. Available from: http://www.law.du.edu/documents/denver-university-law-review/v86_i1_brody.pdf [Accessed 3 December 2014].
- Colbran, S., 2001. Judicial Performance Evaluation, Promotion and Salary Packaging. *The Judicial Review*, 5 (4), 335-357.
- Colbran, S., 2002a. Judicial Performance Evaluation and Settlement Skills. *Journal of Judicial Administration*, 11 (4), 180-204.
- Colbran, S., 2002b. Judicial Performance Evaluation: Accountability without Compliance. *Australian Law Journal*, 76 (4), 235-249.
- Colbran, S., 2003a. Independence and Integrity as a Criterion for Judicial Performance Evaluation. *University of Notre Dame (Australia) Law Review* [online], 5, 15-36. Available from: <http://www.austlii.edu.au/au/journals/UNDAULawRw/2003/3.html> [Accessed 3 December 2014].
- Colbran, S., 2003b. The Limits of Judicial Accountability: The Role of Judicial Performance Evaluation. *Legal Ethics*, 6 (1), 55-72.
- Contini, F. and Mohr, R., 2007. Reconciling Independence and Accountability in Judicial Systems. *Utrecht Law Journal* [online], 3 (2), 26-43. Available from:

- <http://www.utrechtlawreview.org/index.php/ulr/article/view/46> [Accessed 3 December 2014].
- Contini, F., Mohr, R. and 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> [Accessed 24 December 2014].
- Drummond, D., 2001a. Towards a More Compliant Judiciary? - Part I. *Australian Law Journal*, 75 (5), 304-325.
- Drummond, D., 2001b. Towards a More Compliant Judiciary? - Part II. *Australian Law Journal*, 75 (6), 356-377.
- Elek, J.K., Rottman, D.B. and Cutler, B.L., 2012. Judicial performance evaluation: Steps to improve survey process and management. *Judicature*, 96, 65-75.
- Felter, E., 2008. Accountability in the Administrative Law Judiciary: The Right and the Wrong Kind. *Denver University Law Review* [online], 86 (1), 157-193. Available from: http://www.law.du.edu/documents/denver-university-law-review/v86_i1_felter.pdf [Accessed 3 December 2014].
- Gething, M., 2008a. A Pathway to Excellence for a Court – Part I: Defining the Pathway. *Journal of Judicial Administration*, 17 (4), 237-252.
- Gething, M., 2008b. A Pathway to Excellence for a Court – Part II: Defining Excellence. *Journal of Judicial Administration*, 18 (1), 22-38.
- Geyh, C., 2003. Why Judicial Elections Stink. *Ohio State Law Journal* [online], 64 (1), 43-79. Available from: http://moritzlaw.osu.edu/students/groups/oslj/files/2012/03/64.1.geyh_.pdf [Accessed 3 December 2014].
- Geyh, C., 2008. Straddling the Fence Between Truth and Pretense: The Role of Law and Preference in Judicial Decision-Making and the Future of Judicial Independence. *Notre Dame Journal of Law, Ethics and Public Policy* [online], 22 (2), 435-450. Available from: <http://ssrn.com/abstract=1497004> [Accessed 3 December 2014].
- Gill, R., 2012. *Judicial Performance Evaluations as Biased and Invalid Measures: Why the ABA Guidelines Are Not Good Enough* [online]. Available from: <http://ssrn.com/abstract=2031800> [Accessed 3 December 2014].
- Gleeson, A., 1979. Judging the Judges. *Australian Law Journal*, 53 (7), 338-347.
- Goldring, J., 2003. The Reflective Judge: Aims and Content of Judicial Education. *Journal of Judicial Administration*, 12 (4), 216-224.
- Goodman, S., 1982. Judicial Selection and the Qualities that Make a 'Good' Judge'. *The ANNALS of the American Academy of Political and Social Science*, 462 (1), 112-124.
- Griffin, J., 1995. From the Bench: Judging the Judges. *Litigation*, 21 (3), 5-7, 60-62
- Haines, C., 2010. Judicial Independence and Judicial Accountability: How Judicial Evaluations can Support and Enhance Both. *Duquesne Law Review*, 48 (4), 909-928.
- Handsley, E., 2001a. Can Public Sector Approaches to Accountability be Applied to the Judiciary. *Law in Context*, 18 (1), 62-101.
- Handsley, E., 2001b. Issues Paper on Judicial Accountability. *Journal of Judicial Administration*, 10 (4), 180-226.

- Jayawickrama, N., 2002. Developing a Concept of Judicial Accountability – The Judicial Integrity Group and the Bangalore Principles of Judicial Conduct. *Commonwealth Law Bulletin*, 28 (2) 1091-1108.
- Kirby, M., 2003. Judicial Accountability in Australia. *Legal Ethics*, 6 (1), 41-54.
- Kitto, F., 1992. Why Write Judgments? *Australian Law Journal*, 66 (12), 787-799.
- Koebel, T., 1983. Evaluating Judges: An Analysis of Survey Results. *Evaluation Review*, 7 (5), 659-684.
- Kosar, D., 2013. The Least Accountable Branch. *International Journal of Constitutional Law*, 11 (1), 234-260.
- Kourlis, R. and Singer, J., 2007. Using Judicial Performance Evaluations to Promote Judicial Accountability. *Judicature* [online], 90 (5), 200-207. Available from: http://www.americanbar.org/content/dam/aba/migrated/jd/lawyersconf/performanceresource/article/200_207_Kourlis_905_authcheckdam.pdf [Accessed 3 December 2014].
- Kourlis, R. and Singer, J., 2008. A Performance Evaluation Program for the Federal Judiciary. *Denver Law Journal* [online], 86 (1), 7-51. Available from: http://www.law.du.edu/documents/denver-university-law-review/v86_i1_kourlis_singer.pdf [Accessed 3 December 2014].
- Kourlis, R. and Singer, J., 2009. A Strategy for Judicial Performance Evaluation in New York. *Albany Law Review* [online], 72 (3), 655-671. Available from: http://www.albanylawreview.org/Articles/Vol72_3/72.3.0007%20Kourlis.pdf [Accessed 3 December 2014].
- Le Sueur, A., 2004. Developing Mechanisms for Judicial Accountability in the UK. *Legal Studies*, 24 (1-2), 73-98.
- McIntyre, J., 2013. *The Nature and Implications of the Judicial Function*. Thesis (PhD). University of Cambridge.
- Mohr, R. and Contini, F., 2007. Judicial Evaluation in Context: Principles, Practices and Promises in Nine European Countries. *European Journal of Law and Society* [online], 1 (2), 1-40. Available from: <http://www.ejls.eu/2/30UK.pdf> [Accessed 3 December 2014].
- Mohr, R., et al., 1996. Performance Measurement for Australian Courts. *Journal of Judicial Administration*, 6 (3), 156-170.
- Musolf, L., 1998. Performance Evaluation of Federal Administrative Law Judges: Challenge for Public Administration? *The American Review of Public Administration*, 28 (4), 390-401.
- Ng, G. Y., 2007. *Quality of Judicial Organisation and Checks and Balances* [online]. Thesis (PhD). University Utrecht. Available from: <http://dspace.library.uu.nl/bitstream/handle/1874/21459/full.pdf> [Accessed 3 December 2014].
- O'Connor, S.D., 2008. Judicial Accountability Must Safeguard, Not Threaten, Judicial Independence: An Introduction. *Denver University Law Review* [online], 86 (1), 1-6. Available from: http://www.law.du.edu/documents/denver-university-law-review/v86_i1_oconnor.pdf [Accessed 3 December 2014].
- O'Neil, O., 2002. *A Question of Trust: Reith Lectures 2002 (Lecture 1)* [online]. Available from: http://downloads.bbc.co.uk/rmhttp/radio4/transcripts/20020403_reith.pdf [Accessed 3 December 2014].
- O'Ryan, S. and Lansdell, T., 2000. Benchmarking and Productivity for the Judiciary. *Journal of Judicial Administration*, 10 (1), 25-49.

- Paynter, S. and Kearney, R., 2010. Who Watches the Watchmen?: Evaluating Judicial Performance in the American States. *Administration & Society*, 41 (8), 923-953.
- Pelander, J., 1998. Judicial Performance Review in Arizona: Goals, Practical Effects and Concerns. *Arizona State Law Journal*, 30 (3), 643-759.
- Pimentel, D., 2009. Reframing the Independence v Accountability Debate: Defining Judicial Structure in Light of Judge's Courage and Integrity, *Cleveland State Law Review* [online], 57 (1), 1-33. Available from: <http://engagedscholarship.csuohio.edu/clevstlrev/vol57/iss1/3> [Accessed 3 December 2014].
- R v Sussex Justices* [1924] 1 KB 256.
- Rares, S., 2011. What is a Quality Judiciary? *Journal of Judicial Administration*, 20 (2), 133-145.
- Riddel, T., Hausegger, L. and Hennigar, M., 2012. Evaluating Federally Appointed Judges in Canada: Analysing the Controversy. *Osgood Hall Law Journal* [online], 50 (2), 403-427. Available from: <http://digitalcommons.osgoode.yorku.ca/ohlj/vol50/iss2/3/> [Accessed 3 December 2014].
- Riedel, J., 2014. Individual Evaluation of Judges in Germany. *Oñati Socio-legal Series* [online], 4 (5), 974-992. Available from: <http://ssrn.com/abstract=2533857> [Accessed 23 December 2014].
- Roach Anleu, S. and Mack, K., 2014. Judicial Performance and Experiences of Judicial Work: Findings from Socio-Legal Research. *Oñati Socio-legal Series* [online], 4 (5), 1015-1040. Available from: <http://ssrn.com/abstract=2533861> [Accessed 23 December 2014].
- Sackville, R., 2002. From Access to Justice to Managing Justice: The Transformation of the Judicial Role. *Journal of Judicial Administration*, 12 (1), 5-24.
- Seibert-Fohr, A., 2010. *European Comparative Perspectives on the Rule of Law and Independent Courts*. Heidelberg: Max Plank Institute for Comparative Public Law and International Law.
- Soeharno, J., 2009. *The Integrity of the Judge: A Philosophical Inquiry*. Surrey: Ashgate.
- Spigelman, J., 2000a. Seen to be Done: The Principle of Open Justice – Part I. *Australian Law Journal*, 74, 290-297.
- Spigelman, J., 2000b. Seen to be Done: The Principle of Open Justice – Part II. *Australian Law Journal*, 74, 378-383.
- Spigelman, J., 2002. Judicial Accountability and Performance Indicators. *Civil Justice Quarterly*, 21, 18-28.
- Spigelman, J., 2006. Measuring Court Performance. *Journal of Judicial Administration*, 16 (2), 69-80.
- Sponzo, M., 1987. Independence vs Accountability: Connecticut's Judicial Evaluation Program Has Independence and Accountability on the Same Side. *The Judges' Journal*, 26 (2), 13-43.
- Tarr, G., 2009. Do Retention Elections Work? *Missouri Law Review* [online], 74 (3), 605-633. Available from: <http://scholarship.law.missouri.edu/mlr/vol74/iss3/10> [Accessed 3 December 2014].
- Tinkham, T., 2011. Applying a Rational Approach to Judicial Independence and Accountability on Contemporary Issues. *William Mitchell Law Review* [online],

- 37 (3), 1633-1665. Available from:
<http://open.wmitchell.edu/wmlr/vol37/iss3/4> [Accessed 3 December 2014].
- UK Judicial Appointment Commission, s.d. *Qualities and abilities* [online]. London: Judicial Appointment Commission. Available from:
<http://jac.judiciary.gov.uk/application-process/qualities-and-abilities.htm>
[Accessed 22 December 2014].
- Warren, M., 2011. Enhancing our Self-perception: 360-degree Feedback for Judicial Officers. *Journal of Judicial Administration*, 21 (1), 3-7.
- Weatherburn, D., 1996. Performance Indicators for Case Processing in Trial Courts. *Journal of Judicial Administration*, 6 (2), 113-131.
- White, P., 2002. Judging Judges: Securing Judicial Independence by Use of Judicial Performance Evaluation. *Fordham Urban Law Journal* [online], 29 (3), 1053-1077. Available from: <http://ir.lawnet.fordham.edu/ulj/vol29/iss3/11/> [Accessed 3 December 2014].
- White, P., 2009. Using Judicial Performance Evaluations to Supplement Inappropriate Voter Cues and Enhance Judicial Legitimacy. *Missouri Law Review* [online], 74 (3), 635-666. Available from:
<http://scholarship.law.missouri.edu/mlr/vol74/iss3/11> [Accessed 3 December 2014].
- Woodhouse, D., 2005. Judges and the Lord Chancellor in the changing United Kingdom Constitution: Independence and accountability. *Public Law Review*, 16, 227-244.