

Temperament as a Criterion for Judicial Performance Evaluation

STEPHEN COLBRAN*

This article proposes six measures of judicial temperament that may be used in an Australian pilot program for judicial performance evaluation. The measures are evaluated by a national survey of barristers' temperament ratings of Supreme and Federal Court judges. Barristers who responded to the survey thought that the proposed measures of judicial temperament were important measures of judicial performance. The measures have a high degree of internal consistency. The article also examines aspects of judicial temperament as a measure of judicial performance. These aspects include: identifying potential measures of judicial temperament; whether judicial temperament differs between trial and appellate judges; whether judicial gender affects temperament; whether judicial temperament deteriorates with age or experience; whether older judges treat junior barristers differently.

There is no doubt that the manner in which judges conduct themselves is an essential part of justice.¹ The image of a judge is important – a positive image creates respect for the judiciary² and the rule of law. An important aspect of a positive image is judicial temperament. The demeanour and behaviour of judges can affect the performance of barristers in the presentation of their case, of witnesses, and the views of a jury.³ Wood suggests that:

* BCom(Hons), LLB(Hons), LLM(Hons), PhD; Senior Lecturer, Faculty of Law, Queensland University of Technology. The Australian Law Council Foundation funded this research. The author acknowledges the valuable assistance of Associate Professor Art Shulman, Graduate School of Management, University of Queensland.

¹ J Doyle, 'Judgment Writing: Are There Needs for Change?' (1999) 73 *Australian Law Journal* 737.

² American Bar Association, *American Bar Association Guidelines for the Evaluation of Judicial Performance* (1985) 15.

³ Australian Law Reform Commission, *Equality Before the Law*, Discussion Paper No 54 (1993) 26.

No one would question that they [judges] should be civil, courteous and reasonably patient towards all persons in their courts, including parties and their legal representatives, witnesses, spectators and court staff.⁴

Eugene Ehrlich goes so far as to suggest that there is 'no guarantee of justice except the personality of the judge'.⁵

Given that temperament is an important aspect of the judicial role, how do we begin to define its boundaries? Michael O'Mahony⁶ suggests a satirical taxonomy of judicial temperament based on the observations of Rhadamathus. There are nine judicial stereotypes:

- the gentle judge;
- the quiet judge;
- the pragmatic judge;
- the witty judge;
- the lawyer judge;
- the intrusive judge;
- the impatient judge;
- the authoritarian judge; and
- the intellectually challenged judge.

Barristers will no doubt recognise these personality stereotypes and perhaps those judges to whom they relate. The stereotypes do not offer viable measures of judicial temperament.

Judicial performance evaluation directed to judicial temperament can identify problems, which may be dealt with by judicial self-improvement, peer pressure, intervention by the Chief Justice, or in rare cases, disciplinary procedures.

The importance of judicial temperament is also recognised across the Pacific. Temperament is used as a performance evaluation criterion in Alaska, New Jersey, Hawaii, Arizona, Nova Scotia, and appears in the *American Bar Association Guidelines for Judicial Performance Evaluation*. Temperament is also reflected in American judicial codes of conduct.⁷

In Australia, Chief Justice Gleeson suggests that:

⁴ D Wood, *Judicial Ethics: A Discussion Paper* (1996) 15.

⁵ E Ehrlich, 'Freedom of Decision' (1917) 9 *Modern Legal Philosophy Series* 65.

⁶ M O'Mahoney, 'An Irish Litigation Solicitor's Perspective on the Judiciary' (Paper presented at the 18th Australian Institute of Judicial Administration Annual Conference, Darwin, 14-16 July 2000).

⁷ See, for example, the Wyoming *Code of Judicial Conduct*, canon 3 (1994).

modern lawyers, litigants and witnesses, and the public generally, are much more ready to criticise judges whose behaviour departs from appropriate guidelines of civility and judicial attachment. This is a good thing. If judges behave inappropriately, they should be criticised.⁸

Criticism is one thing, but systematic performance evaluation quite another.

Measures of Judicial Temperament

The American studies of Rosenberg,⁹ Watson and Downing,¹⁰ and Judge Shientag¹¹ all point to the importance of the personality factor or the concept of judicial temperament.¹²

Several American States and Nova Scotia include measures of judicial temperament in judicial performance evaluation programs. These programs have various aims, ranging from voter advice in judicial retention elections through to strategies to assist with judicial self-improvement. What is constant amongst the approaches is that they generally measure some aspects of courtroom demeanour (for example, courtesy, control and attentiveness).

'Alaska pioneered the concept of judicial performance evaluation by adopting a statutory judicial evaluation program in 1975.'¹³ Alaska conducts judicial performance evaluation with a primary focus on informing citizens about applicants for judicial retention elections. Judicial self-improvement is of secondary concern, though this is becoming increasingly important. The Alaskan program measures judicial temperament by having attorneys rate judicial courtesy, freedom from arrogance, human understanding and compassion, and ability to control the courtroom.

⁸ A Gleeson, 'Performing the Role of the Judge' (1988) 10(8) *Judicial Officers' Bulletin* 57, 58.

⁹ M Rosenberg, 'The Qualities of Justice: Are They Strainable?' (1966) 44 *Texas Law Review* 1063.

¹⁰ R Watson and R Downing, *The Politics of the Bench and the Bar: Judicial Selection Under the Missouri Non-Partisan Court Plan* (1969) 293.

¹¹ B Shientag, *The Personality of the Judges* (1944).

¹² American Judicature Society, *Handbook for Judicial Nominating Commissions* (1984) 57-63.

¹³ J Pelander, 'Judicial Performance Review in Arizona: Goals, Practical Effects and Concerns' (1998) 30 *Arizona State Law Journal* 643, 651; ALASKA STAT § 15.58.050 (Michie 1996); Alaska Admin R, 23(c).

New Jersey was the second American State to explore judicial performance evaluation.¹⁴ Unlike the first experiments in Alaska, which focused on judicial retention elections, the New Jersey program focused on judicial self-improvement. This is an approach directly relevant in the Australian context. The New Jersey program is a well established and widely respected program often copied to varying degrees by other jurisdictions, for example, Hawaii and Nova Scotia.

The New Jersey program measures temperament under the heading 'comportment'. Measures focus on a judge's attentiveness, courtesy, open-mindedness, patience, absence of arrogance, listening skills, decisiveness, even-handed treatment of attorneys, fostering a general sense of fairness, and the absence of bias based on race, gender, ethnicity, religion or social class. The New Jersey program includes under the umbrella of 'temperament' aspects of relevant knowledge and decision-making. The measures proposed by this study do not include substantive measures of this nature.

The Hawaii program has exactly the same measures as New Jersey, except that they also consider even-handed treatment of litigants. Nova Scotia is also similar to New Jersey, but also includes measures such as courtesy to staff, dignity, and sensitivity to the impact of his or her demeanour. The New Jersey, Hawaii and Nova Scotia programs are largely based on attorney surveys.

The Arizona program has more extensive data sources. Questions are asked of attorneys, litigants, witnesses, self-litigants, jurors and court staff concerning judicial temperament, based on measures including a judge's understanding and compassion, dignified demeanour, courteous conduct that promoted public confidence in the court, and the judge's ability.

In 1985, the American Bar Association adopted the *Guidelines for the Evaluation of Judicial Performance*. The Guidelines contain judicial performance evaluation criteria, methodological and administrative guidelines, and proposals on the use and dissemination of results. Temperament is included in the guidelines under the heading 'prepa-

¹⁴ In the early 1980s the Supreme Court established a permanent committee on judicial evaluation and performance. Members included lawyers and judges. Chief or presiding judges evaluated other judges. The focus was on competence, productivity and conduct. A consultant from the Graduate School of Management, Rutgers University, was used to determine suitable methodological approaches. Four years of planning resulted in a pilot program. The pilot program ran from 1983 to 1987. The program became permanent in 1988.

ration, attentiveness, and control over proceedings', and is measured by:

- courtesy to all parties and participants; and
- willingness to permit every person legally interested in a proceeding to be heard, unless precluded by law or rules of court.

A further characteristic suggested by the Australian Bar Association is that judges may need to demonstrate an ability to understand community values and be prepared to apply them.¹⁵ They also mention practicality, common sense and a well-developed sense of the role of the courts and the law. This has sometimes been perceived as 'breadth of vision'.¹⁶ The judicial temperament may also include 'practicality and common sense,¹⁷ vision'.¹⁸ All such personal judicial qualities are important, but are very difficult to measure, and are for that reason not included in the measures proposed by this study.

Sir Harry Gibbs mentions other aspects of judicial temperament in the context of judgment writing. His Honour observed:

Although the judge should, if it is relevant, make it clear that he or she rejects the evidence of a particular witness, he or she should refrain from censoring or condemning the witness – for example calling a witness a liar or a rogue – unless it is strictly necessary for the purpose of the judgement to do so. The judge is there to decide the case rather than to denounce human evil or folly. I digress to add that similarly a judge should avoid making derogatory remarks (at least publicly) about another judge from whom an appeal has been brought, or about counsel, however much the latter may have tried the judge's patience.¹⁹

¹⁵ Australian Bar Association, submission in response to the Attorney-General's Discussion Paper, *Judicial Appointments – Procedure and Criteria*, 1 December 1993 (1993) 5.

¹⁶ Z Cowan and K Ryan, submission to *Report of the Select Committee of the New South Wales Legislative Assembly on the Appointment of Judges to the High Court of Australia* (1975); Australian Bar Association, above n 15.

¹⁷ W Wells, *Law, Judges and Justice* (1991) 68.

¹⁸ Law Council of Australia, *Judicial Appointments Procedure and Criteria* (1993) 6, 8, 12, 93.

¹⁹ H Gibbs, 'Judgment Writing' (1993) 67 *Australian Law Journal* 494, 497-8. As Francis Bacon said (Francis Bacon, *Of Judicature* (1612)), patience is 'an essential part of justice'.

Table 1: Hypotheses

Hypothesis 1	Judicial temperament can be reliably measured by using survey instruments with barristers and with judicial officers
Hypothesis 2	Permanent appellate judges will have higher temperament performance ratings than first instance judges
Hypothesis 3	Female judges will have lower temperament performance ratings than male judges
Hypothesis 4	Judges 65 years of age or older will have lower temperament performance ratings than younger judges
Hypothesis 5	Judges in their first five years of office will have higher temperament performance ratings than more senior judges
Hypothesis 6	Experienced barristers give higher temperament performance ratings than inexperienced barristers regardless of the judge
Hypothesis 7	Younger barristers will give lower judicial temperament ratings than more senior barristers

In adopting a more limited view than the United States and Canadian approaches, six measures of judicial temperament are proposed for the purpose of testing the hypotheses in Table 1:

- courtesy, freedom from arrogance;
- human understanding and compassion;
- displays patience;
- ability to control the courtroom;
- promotes public confidence; and
- dignified demeanour.

The proposed measures of judicial temperament do not probe into the substantive personality of the judge, the fairness of the decisions made, or the judge's understanding of law or community values. Any participant in the litigation – judges, barristers, jurors, witnesses, litigants, court staff and court watchers, can assess temperament. The data sources used in this article are derived from a national survey of barristers.

Does Temperament Differ Between Trial and Permanent Appellate Judges?

Sterling, Stott and Weller²⁰ argue that separate questionnaires should be developed for appellate and trial judges to reflect their unique characteristics.²¹ However, this may not be relevant to judicial tem-

²⁰ J Sterling, K Stott and S Weller, 'What Judges Think of Performance Evaluation: A Report on the Colorado Survey' (1981) 64(9) *Judicature* 414.

²¹ See also R Hanson, 'Appellate Court Performance Standards', submitted to the State Justice Institute by the National Center for State Courts and the Appellate Court Performance Guidelines Commission (1995).

perament. Judicial temperament is equally important for both trial and appellate judges. For example, 'Patience and courtesy are essential qualities for all judges, but particularly for a trial judge. The failure to display these qualities can reflect adversely on an otherwise able judge.'²² The American Bar Association argues:

While appropriate courtroom demeanour is possibly more important for a trial judge – in view of the trial judge's frequent interaction with the public – it is also an important criterion for an appellate judge. This criterion asks the evaluator to measure how well the judge listens and whether the judge is fair and courteous to counsel. At the appellate level, this criterion is, for the most part, limited to oral argument.²³

It is possible to compare the temperament ratings given by barristers to appellate judges with trial judges in those jurisdictions with full-time appellate judges, for example, Queensland, New South Wales and Victoria. This is the subject of hypothesis 2.

Are Judicial Temperament Ratings Affected by Gender?

An interesting question is whether or not there is a difference on the judicial temperament scale between men and women judges. 'Men and women have different perceptions of human relationships and of society, [such] perceptions have an influence on judicial decision making in general.'²⁴ On this view, women judges will have judicial temperament scores different than men judges, because women judges may differ in their judicial decision-making from men judges. This is an argument for greater female judicial representation. If this is the case, women judges should perhaps be evaluated differently from male counterparts, with criteria and measures sensitive to gender issues.

An alternative view arises with studies that suggest that precisely the same task is evaluated differently depending on whether it is performed by a man or a woman (for example, the same paper read to different audiences by men and women is likely to be assessed overall as more scholarly when read by a man). Such bias may also be evident in assessments of judicial temperament. In other words, women judges will have judicial temperament scores different than men judges, because barristers, with bias, will give different judicial tem-

²² American Bar Association, above n 2, 15.

²³ *Ibid.*

²⁴ D Malcolm, *Report of Chief Justice's Task Force on Gender Bias* (1994) 90.

perament scores to women judges, even though their behaviour as judges is the same as their male counterparts.

These assertions are the subject of hypothesis 3, which questions whether gender has any association with judicial temperament ratings measured by this study.

Does Judicial Temperament Deteriorate with Age or Experience?

The Constitutions of the Commonwealth and the States require that judges retire at the age of 70²⁵ or 72. The rationale is that they are too old to suitably carry out their duties at the required standard. There has also been a trend towards early retirement from the bench. It is unclear whether the reason is stress related burnout or economic factors.²⁶

Justice Thomas, when referring to judicial stress, said:

You may feel an excitement in the lower intestine as you prepare to walk into court. The reason is that you are expected to perform. It gets worse as you get older. It is so easy to lose whatever reputation you have built up through one silly statement. And there is constantly that pressure to get it right. You need adrenalin or pressure, to produce your best work.²⁷

The effects of judicial age and experience on temperament are the subject of hypotheses 4 and 5.

Do Older Judges Treat Junior Barristers Differently?

Anecdotal evidence would suggest that judges tend to give junior barristers a harder time in court than more senior barristers. Some barristers suggest frequent and detailed probing of legal principles. If this were the case, it might be expected that junior barristers as a group would give lower ratings of judicial temperament than more senior barristers. Other reasons could also account for any observed difference. Possibly, barristers might prefer the temperament of judges who are more like them in terms of age or experience. The possibility of observable differences is the subject of hypotheses 6 and 7.

²⁵ *Judges' Retirement Act of 1921*, 12 Geo V No 14 (Qld) s 3; *Commonwealth Constitution* s 72; *Judicial Officers Act 1986* (NSW) s 44(1).

²⁶ P Young, 'Judges' Retirements' (1997) 71 *Australian Law Journal* 733.

²⁷ J Thomas, 'Get Up Off the Ground: A Commentary on Hon Kirby J's "Judicial Stress – An Update"' (1997) 71 *Australian Law Journal* 785, 787.

Survey Instrument and Sampling Procedures

Australian barristers were asked to evaluate the temperament of sitting Supreme and Federal Court judges using the six measures of judicial temperament previously developed.

The survey instrument was constructed to gather barristers' general impressions concerning performance evaluation before they attempted a structured analysis of the temperament of sitting Supreme and Federal Court judges.²⁸

Survey booklets containing the names of all sitting Supreme and Federal Court judges were distributed to the population of 4218 practicing barristers listed in the Law Council of Australia's *Australian Legal Directory* 1999 edition, double checked against the Yellow Pages Online. No follow-up survey instrument was used.²⁹

By the final cut-off date of 21 December 1999, a total of 270 survey booklets were returned.³⁰ The overall response rate for barristers was 6.40 per cent. The low response rate raises questions as to the representativeness of the sample. Comparisons with known population statistics indicated no statistically significant bias based on jurisdiction or gender, except for Victoria and Western Australia. In these jurisdictions, relatively more female barristers responded to the survey than would be expected from the population.

There is no easy way of determining whether non-response was due to lack of knowledge about the judges concerned or for other reasons. The results of this study are presented as that of the survey respondents only.

²⁸ The instrument was pre-tested with thirty experienced barristers before being finalised. Ten jurisdiction-specific survey booklets were created. One survey instrument was created for each State and Territory, except New South Wales and Victoria. Two survey instruments were created for New South Wales and Victoria due to the large number of superior court judges in those jurisdictions. Each survey instrument included a separate document containing an alphabetical list of no more than 32 judge names. The remaining survey instruments are available from the author. The survey instrument was written in plain English.

²⁹ The potential for bias from barristers with an axe to grind against a particular judge or court, duplicating low ratings, presented an unacceptable risk. Ethics requirements precluded identification of barristers who had completed a survey booklet.

³⁰ A random sample of five per cent of the barristers' survey booklets were re-examined to determine the accuracy of data entry. Frequencies of values for each variable were checked for outliers and data entry errors. The initial mailing or delivery to barristers occurred on 13 September 1999. In each case, a self-addressed reply paid return envelope was enclosed. The data was collected over a stated time period (13.9.99 – 21.12.99) rather than on a case-specific basis.

Barristers who appear as advocates before the court are most knowledgeable about judicial performance. They are an appropriate and reliable source of data. The barrister, more than anyone else, has repeated opportunities over extensive periods to view different judicial behaviour, in different contexts, and to compare them. Barristers can make judgements and comparisons with an educated appreciation of how the judicial system works in actual practice, in the context of the cases in which they appear. Barristers' research, writing and oral skills are very similar to those used by judges. This fact, combined with their experience with the judicial function, makes them the cohort of individuals from which superior court judges are appointed.³¹

It is useful to consider those who did not answer the survey instrument. Six per cent of participants in the 0-5 years range of experience completed the survey. This was to be expected, since barristers with little experience are unlikely to frequently practice in superior courts such as the Supreme Court or the Federal Court, the subjects of this study.

Hypothesis 1

Hypothesis 1 addresses the question of whether reliable indices of the core construct, judicial temperament, can be measured. It is different from the remaining hypotheses, as it has a distinct methodological basis. The literature on judicial performance evaluation presents a diversity of views as to whether reliable (in this case, agreement across items that are supposed to measure the same thing) measures can be formed to reflect aspects of judicial performance.

Barristers were asked to rate each measure of temperament on a four-point scale from '1 – very unimportant' to '4 – very important'. A fifth category of '5 – don't know' was included on the scale. The data was collapsed into absolute values of important or unimportant, with 'don't know' and missing responses reported together as non-responses. The results appear in Table 2.

³¹ See P de Jersey, 'The Merit Test' (2000) 4 *Queensland Bar News* 8.

Table 2: Summary of Judicial Temperament Criterion

Measures	Barristers Survey			
	N	Important	Unimportant	Non-responses
Courtesy, freedom from arrogance	113	107 (94.7%)	6 (5.3%)	157
Human understanding and compassion	116	107 (92.2%)	9 (7.8%)	154
Displays patience	115	105 (91.3%)	10 (8.7%)	155
Ability to control court room	114	106 (93%)	8 (7%)	156
Promotes public confidence	113	99 (87.6%)	14 (12.4%)	157
Dignified demeanour	114	90 (78.9%)	24 (21.1%)	156

$p < .05$, two tailed, $df = 1$

Table 2 indicates that all measures of temperament were uniformly and overwhelmingly regarded as important measures of judicial performance. The Alpha (Cronbach) model of internal consistency, based on average inter-item correlation, returned a result of .93, out of a maximum of 1, which suggests a high degree of consistency between the measures of judicial temperament. The data is consistent with acceptance of hypothesis 1.

Hypotheses 2 – 7

The data for hypotheses 2-7 were derived from question 11 of the barristers survey, which states:

In this question you will be asked to rate the performance of sitting Supreme and Federal Court judges based on criteria developed by the American Bar Association. The names of the judges are listed on the accompanying Judicial Names Legend. Please only rate the performance of judges with whom you have had actual court experience in the period January 1997 – August 1999, not merely by reputation.³² Place an 'x' in the box beneath the names of those judges with whom you have had no direct experience during this period then leave the column blank.

If you do not have sufficient personal experience to rate a given characteristic of a particular judge, place an 'x' in the row for that characteristic.

³² Information as to the actual experience before each judge was not sought. Pilot surveys indicated that barristers were unlikely to keep or access such records.

Please rate each judicial performance characteristic according to the following five point 'acceptance scale'.³³

1. Unacceptable Seldom meets minimum standards of performance
2. Deficient Does not always meet minimum standards of performance
3. Acceptable Meets minimum standards of performance
4. Good Often exceeds minimum standards
5. Excellent Consistently exceeds minimum standards

Please write a score out of 5 in the column beneath the name of each judge on the row for each of the stated performance characteristics.

An accompanying 'Judicial Names Legend' stated the names of each judge within each jurisdiction in alphabetic order.

Table 3

Hypotheses	Mean	F	Numerator df	Denominator df	Significance
<i>Hypothesis 2</i> Non-appellate Appellate	3.82 3.86	.77	1	2214	.379
<i>Hypothesis 3</i> Male Female	3.82 3.86	.33	1	2214	.569
<i>Hypothesis 4</i> Judges <65 Judges >=65	3.89 3.59	29.8	1	2214	.000***
<i>Hypothesis 5</i> Judge experience <5 Judge experience >=5	3.96 3.76	19.5	1	2214	.000***
<i>Hypotheses 6 & 7</i> Barrister experience 1-10 10-18 18-40	3.72 3.80 3.86	3.05	1	2214	.048*

Appellate Versus First Instance Judges

Hypothesis 2 states 'Permanent appellate judges will have higher temperament performance ratings than first instance judges'. A multi-variate analysis of variance was used to examine if significant statisti-

³³ This replicates the scale used by the Alaska Judicial Council: <<http://www.ajc.state.ak.us/Retention98/retgen1.htm>> at 23 November 2000.

cal differences existed in how barristers rated appellate judges³⁴ versus non-appellate judges³⁵ on temperament, while co-varying out the effects of barrister jurisdiction and experience.

The results in Table 3 indicate that appellate judges were not rated significantly different from trial judges on temperament.³⁶ Both groups were performing at the high end of the range 'acceptable through good'.

Female Versus Male Judges

Hypothesis 3 states, 'Female judges will be rated significantly lower on the temperament performance criteria compared with male judges'. A multi-variate analysis of variance was used to examine if significant statistical differences existed between male and female judges on each temperament measure, while co-varying out the effects of barrister jurisdiction and experience.³⁷

The results in Table 3 indicate no statistically significant difference between male and female judges concerning judicial temperament. Both groups were performing at the high end of the range 'acceptable through good'.

Old Versus Young Judges

Hypothesis 4 states, 'Judges 65 years of age and older will have lower temperament performance ratings than younger judges'. A multi-variate analysis of variance³⁸ was used to examine if significant statistical differences existed in judges 65 years of age or older³⁹ versus judges less than 65 years of age⁴⁰ on the temperament composite, while co-varying out the effects of barrister jurisdiction and experience. The results appear in Table 3. There is a significant judicial age effect. The results suggest that ratings of judicial temperament decline as judges reach 65 years of age.

³⁴ N = 430.

³⁵ N = 1780.

³⁶ Since each barrister only rated judges whom they had appeared before, and barristers often rated more than one judge, the observations are not independent. Analysis of variance that treats the judges as within comparison was performed, albeit with large numbers of missing values. The resulting patterns were the same.

³⁷ Male (N = 2013), female (N = 197).

³⁸ This analysis breached the assumption for independent samples resulting in higher Fs.

³⁹ N = 421.

⁴⁰ N = 1789.

Inexperienced Versus Experienced Judges

Hypothesis 5 states, 'Judges in their first five years of office will have higher temperament performance ratings than more senior judges'. A multi-variate analysis of variance⁴¹ was used to examine if significant statistical differences existed between judges in their first five years of office⁴² versus more senior judges⁴³ on each temperament measure, while co-varying out the effects of barrister jurisdiction and experience. There is a significant judicial experience effect for judicial temperament, which confirms hypothesis 5.

The performance results related to judicial age are contentious, and rightly so. The observed pattern of lower performance results for judges 65 years or older raises many questions, for example, is lengthy experience always beneficial to performance? What impact does stress and age really have on performance? These are questions relevant to the entire population, not just the judiciary. Many differing interpretations follow from the results, for example:

- judges 65 years or older are discriminated against by biased respondents;
- the retirement age of judges should be lowered to 65 to be consistent with the norm society has placed on all other workers;
- elderly judges have the 'right' perspective, and their younger peers and the profession need enlightenment;
- less experienced, often younger judges are more attuned to the needs of the bar;
- younger barristers, more current in their legal training and closer to the norms of current society, may have different views of the law or have higher standards of courtroom performance;
- perhaps younger barristers feel victimised by older judges;
- the workload of judges 65 years or older should be lowered. This is consistent with the supernumerary status of some judges in Canada;⁴⁴ or

⁴¹ This analysis breached the assumption for independent samples resulting in higher Fs.

⁴² N = 697.

⁴³ N = 1547.

⁴⁴ In Canada, federal legislation creates the category of a supernumerary judge, being a federally appointed judge who, having served 15 years on the bench and having attained the age of 65 (whichever last occurs), has the right to elect supernumerary status. This entitles a judge, on full salary and status, to sit about a third of the time, as the Chief Justice may arrange. The legislation applies to the Court of Appeal and Trial Court of Ontario but not the Supreme Court of

- the workload of judges 65 years or older should be lowered and judicial education training courses designed for their specific needs. This recognises the vital contribution elderly judges make, but reduces their workload to take account of factors such as age, stress etc, and caters for educational programmes designed for their specific needs in relation to judicial self-improvement.

There are countless other views that may be argued in response to the observed results. Further research is needed to address why elderly judges adopt the views and approaches they do, and why the profession reacts adversely to them when considering their performance on temperament.

The Effect of Barrister Age and Experience

Hypothesis 6 states, 'Experienced barristers will give higher temperament performance ratings than inexperienced barristers regardless of the judge'. A multi-variate analysis of variance was used to examine if significant statistical differences existed between three levels of barrister experience⁴⁵ on each performance composite, while co-varying out the effects of barrister gender.

Table 3 reveals a significant barrister experience effect for the judicial temperament. Given this analysis breached the assumption for independent samples resulting in higher Fs, the result for judicial temperament must be interpreted carefully. The results tend to confirm hypothesis 6.

Hypothesis 7 states, 'Younger barristers will give lower judicial temperament ratings than more senior barristers'. No data concerning age was collected to protect the anonymity of respondents. Experience may be used as a proxy measure, directly related to age. The results for hypothesis 6 can be used for hypothesis 7. The data revealed a significant positive relationship between both the experience and age of the barrister and the rating on the judicial temperament

Canada: W Estey, 'The North American Experience: A Theorem on Judicial Administration' in Victoria Law Foundation, *Seminar on Constitutional and Administrative Responsibilities for the Administration of Justice: The Partnership of Judiciary and Executive* (1985) 35. This approach implicitly recognises the reduced workload capacities of such judges due to age, while maintaining the useful contribution of the experienced judiciary. The other reason for this practice is the recognition of inadequate pension schemes. The latter reason is not relevant to Australia.

⁴⁵ 1-10 years N = 306, 10-18 years N = 654, 18-40 years N = 1250.

composite.⁴⁶ Hypothesis 7 is confirmed based on the results of barristers who completed the survey.

Anecdotally, the older and more experienced the judge, the lower the temperament ratings from younger practitioners. Two separate univariate analyses of variance were conducted for the dependant variable, judicial temperament. The fixed factors were judicial experience (0-5 years, 5-10 years, 10+ years)⁴⁷ and judicial age (40-50 years, 50-60 years, 60+ years) respectively.⁴⁸ Barristers' experience and sex were removed (co-varied out). In both analyses there was a main effect for judicial experience⁴⁹ and judicial age.⁵⁰ There were no other significant effects. This result suggests that older and more experienced judges receive lower ratings from barristers despite the barristers' experience or sex. This may be a reflection of Justice Young's view that older judges may be perceived as 'grumpy old men'.⁵¹ It seems the notion that barristers might prefer the temperament of judges who are more like them in terms of age and experience is not supported by the results.

Conclusion

Barristers who answered the survey instrument uniformly and overwhelmingly thought that the following measures of judicial temperament were important measures of judicial performance:

- courtesy, freedom from arrogance;
- human understanding and compassion;
- displays patience;
- ability to control court room;
- promotes public confidence;
- dignified demeanour.

The internal consistency of the measures was particularly strong. These measures can therefore be reliably used with barristers in the context of evaluating the performance of judges based on temperament.

⁴⁶ A composite of all six measures of judicial temperament.

⁴⁷ N= 687, 653, 870 respectively.

⁴⁸ N = 115, 1004, 1091 respectively.

⁴⁹ Mean 0-5 years = 3.961, 5-10 years = 3.846, 10 years+ = 3.703, F = 13.905, df = 2, Sig. = .000.

⁵⁰ Mean 40-50 years = 3.977, 50-60 years = 3.928, 60 years+ = 3.715, F = 8.877, df = 2, Sig. = .000.

⁵¹ P Young, 'Judge's Popularity Poll' (1999) 73(12) *Australian Law Journal* 855.

Neither the gender of a judge, nor whether they were a permanent appellate judge, had any significant statistical relationship with judicial temperament. The temperament of female and male judges was indistinguishable, as was the temperament of appellate and trial judges.

A statistically significant positive relationship was evident between both the experience and age of the barrister and ratings of judicial temperament. Younger, inexperienced barristers tended to give judges lower temperament ratings. Various alternative explanations may be given for the observed result. Further research is needed to probe this intriguing result.

Judicial temperament ratings also decline with judicial age and experience. Judicial officers 65 years of age and older have statistically significantly lower temperament ratings than younger judges. The corollary that judges in their first five years of office will have higher temperament performance ratings than more senior judges was also affirmed. This was the case even after controlling for barrister gender, jurisdiction, and experience.

Consistent with the approaches adopted in Alaska, New Jersey, Hawaii, Arizona, Nova Scotia, and by the American Bar Association, any pilot judicial performance evaluation program in Australia should include measures related to judicial temperament. Should such a program include barrister surveys, the measures proposed and tested in this article may be of assistance.