

Spring 3-1-2013

Criminal Defense Lawyer Moneyball: A Demonstration Project

Ronald F. Wright

Ralph A. Peebles

Follow this and additional works at: <https://scholarlycommons.law.wlu.edu/wlulr>



Part of the [Constitutional Law Commons](#), [Criminal Law Commons](#), and the [Criminal Procedure Commons](#)

Recommended Citation

Ronald F. Wright and Ralph A. Peebles, *Criminal Defense Lawyer Moneyball: A Demonstration Project*, 70 Wash. & Lee L. Rev. 1221 (2013), <https://scholarlycommons.law.wlu.edu/wlulr/vol70/iss2/12>

This Panel 3: New Cases and New Tactics: Approaching Gideon Through a Modern Lens is brought to you for free and open access by the Washington and Lee Law Review at Washington & Lee University School of Law Scholarly Commons. It has been accepted for inclusion in Washington and Lee Law Review by an authorized editor of Washington & Lee University School of Law Scholarly Commons. For more information, please contact lawref@wlu.edu.

Criminal Defense Lawyer Moneyball: A Demonstration Project

Ronald F. Wright & Ralph A. Peeples*

Table of Contents

I. Introduction	1222
II. Evaluation by Guesswork.....	1224
A. Guesswork by Courtroom Professionals	1225
B. Organizational and Client Guesswork.....	1228
III. Evaluation in a Data-Impoverished Context	1231
IV. A Bottom-Line System to Rate Attorneys.....	1236
A. Comparing Public Defenders to Private Defense Attorneys	1236
B. Attorney Ratings in a Guideline Sentencing World	1241
1. Choice of Attorney Quality Measure.....	1242
2. Starting Point and Value Added	1243
3. Comparisons Among Cases, Crimes, and Districts	1245
4. Compiling the Case Ratings of Individual Attorneys.....	1246
C. Customizing Reports	1250
V. The Environment for Success.....	1251
A. Who Are These Guys?.....	1252

* Ronald Wright is the Needham Y. Gulley Professor of Criminal Law at Wake Forest University; Ralph Peeples is a Professor of Law at Wake Forest University. We are grateful for reality-based observations from Ty Alper, Steve Benjamin, Stephanos Bibas, Josh Bowers, Nora Demleitner, Cara Drinan, Don Dripps, Margareth Etienne, Brandon Garrett, John Gross, Catherine Harris, Erica Hashimoto, J.D. King, Erik Luna, Bob Mosteller, Alexandra Natapoff, Norman Reimer, Jenny Roberts, Abbe Smith, and Robin Steinberg. We would also like to thank Mark Vaders, Tom Watkins, Tito Morales, and Stephanie Beale for excellent research assistance.

B. The Raw Materials and the Attorney Tools.....	1255
VI. The Uses and Users of Performance Scores.....	1262
A. Insider Uses.....	1262
B. Outsider Uses	1264
VII. Conclusion.....	1266

I. Introduction

The book and movie *Moneyball*¹ portray the iconoclastic general manager of a baseball team. When searching for new players, this GM deemphasized the insights of baseball scouts as on-the-scene evaluators of a player's talents, and looked instead to statistical measures of player quality. By using such an atypical evaluation technique, the small-market (and low-budget) team was able to identify high-quality players who escaped the attention of other teams.²

We take this idea from baseball into the criminal courts. In this Article, we pose the question whether criminal defense organizations—either public defender offices or private firms—could meaningfully evaluate the skills of their attorneys through the use of metrics, rather than relying so heavily on the in-person observation of their work in the courtroom. In particular, could managers of criminal defense attorneys assemble performance statistics about their lawyers—numbers that would help them evaluate the attorneys and improve their work?³

1. MICHAEL LEWIS, *MONEYBALL: THE ART OF WINNING AN UNFAIR GAME* (2003); *Moneyball* (Sony Pictures 2011).

2. See *supra* note 1 and accompanying text.

3. See Patrick J. Lamb, *If Legal Services Value Stats Were Created, Standardized, Law Clients Could Play Moneyball*, A.B.A. J. (Sept. 29, 2011, 9:30 AM), http://www.abajournal.com/legalrebels/article/legal_services_value_stats_moneyball/ (last visited Apr. 2, 2013) (on file with the Washington and Lee Law Review). Paul Lippe, *Can Moneyball Principles Be Applied to the Valuation of Legal Services?*, A.B.A. J. (Sept. 29, 2011, 8:29 AM), http://www.abajournal.com/legal_services/article/can_moneyball_principles_be_applied_to_the_valuation_of_legal_services/ (last visited Apr. 2, 2013) (on file with the Washington and Lee Law Review). Both articles discuss the possibility of using statistical analysis to evaluate lawyer performance.

The answer, we believe, is a qualified yes. The data available in most criminal systems make it possible to draw inferences about *groups* of attorneys; limits in the data, however, mean that a reliable, comprehensive rating of *individual* lawyers is not realistic at this point. That being said, a rankings system could offer a *partial* view of individual attorney quality. Numerical ratings might help managers identify some individual attorneys as proper candidates for closer in-person monitoring. If supervisors were to combine statistical ratings with other more individualized audits, they could learn something real and reliable about attorney quality. In short: Yes, it is possible to quantify some meaningful aspects of attorney performance, even in the midst of the messy and underfunded criminal justice systems in the states. Statistical performance-based rankings could support better leadership in defense attorney organizations.

Rather than simply assert that a rating system is possible, we attempt in this paper to show its feasibility. We employ data from the North Carolina courts as a demonstration project to illustrate how an office might develop a rating system for the attorneys who work there. Like the *Moneyball* ratings that focus on a hitter's run-producing results, our attorney ratings are based on the bottom line: sentencing reductions those attorneys achieve for their clients, principally through plea negotiations.

Leaders of criminal defense organizations would not stop with a crude indicator of attorney success; they would want to know how to hire the best attorneys and create a work environment that makes success possible. We therefore use our tentative quality ratings to address the question of structural causes. What makes one attorney noticeably more or less effective than the typical defense lawyer? Is it the attorney's education, or the amount or type of experience she brings to the job? Does race, gender, or age have any bearing? Do the attorney's roots in the local community have any impact? Does the attorney's practice environment, such as the size or type of law office where she practices, make a difference? How about the volume of cases that she handles?

We show here that it is possible to identify some likely causes of defense attorney success. Our most surprising discovery is that experience actually has a negative correlation with performance after the first five years: the more time an attorney has spent in

the profession, the more likely that her clients will obtain a more severe sentence. This finding remains in place even after controlling for the seriousness of the charges and caseloads of the attorneys.

In Part II of this Article, we note the widespread need for careful evaluation of defense attorney quality and the limited tools now available to make that assessment. In Part III we explain that any evaluation of criminal defense attorneys is challenging because these lawyers provide a complex service in a data-poor environment. We also note the efforts in other industries—ranging from sports to education to health care—to use performance statistics as part of an evaluation of complex work. In Part IV, we describe the history of efforts to measure attorney quality through court system statistics and introduce our strategy for measuring attorney quality based on North Carolina data. The rating system that we construct here produces a customized ranking report for a manager of criminal defense attorneys. That report quantifies how much a given attorney affected the sentences for his or her clients, as compared to other attorneys in similar cases. The ranking report also tells the manager the margin of error that attaches to the score for each attorney.

In Part V, we look more closely at the experience and work environment of a subset of North Carolina attorneys in a tentative effort to explain the causes of attorney success. Finally, in Part VI we close with some reflections on other potential users of our statistical rating system, concluding that managers of defense organizations are better situated than judges, prosecutors, or clients to make wise use of comparative ratings.

II. Evaluation by Guesswork

Criminal defense attorneys hold in their hands the liberty, the livelihood, and the very lives of their clients. Any professional who provides services this important must expect to be watched, evaluated, and even second-guessed. Evaluations of defense lawyers do indeed happen from time to time in the criminal

justice system in this country.⁴ That evaluation comes from several different players—both inside and outside the system—including judges, prosecutors, potential clients, and managers of defense lawyer organizations.

Unfortunately, in the typical state court environment, these evaluators do not have the materials they need.⁵ Judges applying the constitutional standards for availability and effectiveness of counsel only ask in gross terms if an attorney is present at all, and if that attorney shows the clearest signs of substandard performance. Judges who appoint defense counsel rely on ad hoc impressions of the quality of attorney performance. Managers of defense lawyer organizations have access to more information about attorney performance, but much of that information goes to waste. And worst of all, potential clients operate almost entirely in the dark about the quality of defense lawyers.

All of these actors find themselves compelled to evaluate criminal defense lawyers, yet they do not find the tools at hand to do the job well. They evaluate by guesswork.

A. Guesswork by Courtroom Professionals

Judges evaluate the work of counsel in at least two settings: when they apply constitutional minimum standards of availability and quality, and when they appoint attorneys for indigent defendants.⁶ In both settings, judges operate on the basis of extremely thin information.

The *Gideon v. Wainwright*⁷ standard determines whether the government has met its obligation to make defense counsel available. In its traditional form, *Gideon* asks a binary question: is a lawyer present during a proceeding when the Constitution

4. See Adele Bernhard, *Raising the Bar: Standards-Based Training, Supervision, and Evaluation*, 75 MO. L. REV. 831, 845–51 (2010) (describing methods for evaluation defense counsel performance).

5. See *id.* at 848 (“[L]ack of oversight is consistently cited as one of the structural defects in indigent defense systems.”).

6. See *infra* notes 7–8, 11–12 and accompanying text.

7. *Gideon v. Wainwright*, 372 U.S. 335 (1963); see also *Alabama v. Shelton*, 535 U.S. 654 (2002).

guarantees the assistance of counsel?⁸ In applying this standard, a judge needs no information about the individual skills of the attorney. Although the information available to the judge is thin, the traditional doctrine requires no effort to compare one attorney to another, or to compare an attorney to an abstract standard of performance.⁹

The availability standard of *Gideon* also takes a more innovative form, leading some courts to ask whether the limited funding and large workloads of defense lawyers create such an overcrowded environment that an attorney is not truly “available”; that is, competent lawyering usually does not happen in such an overburdened system.¹⁰ The judge in this setting does not inquire about the outputs of any single attorney, but hears evidence instead about the resource inputs that are available to entire groups of attorneys.

The ineffective assistance of counsel cases, such as *Strickland v. Washington*,¹¹ require the judge to assemble more evidence about an individual lawyer. In particular, the judge asks whether the attorney performed adequately and whether an

8. See *United States v. Mathur*, 685 F.3d 396, 399 (4th Cir. 2012) (“[T]he right [to counsel] is violated when states fail to appoint counsel for defendants who cannot afford representation.”) (citing *Beard v. Banks*, 542 U.S. 406, 417 (2004)); *State v. Wiley*, 565 S.E.2d 22, 40 (N.C. 2002) (stating that total deprivation of the right to counsel is a structural error); *State v. Anderson*, 558 S.E.2d 87, 92 (N.C. 2002); Justin Marceau & Nathan Rudolph, *The Colorado Counsel Conundrum: Plea Bargaining, Misdemeanors, and the Right to Counsel*, 89 DENV. U. L. REV. 327, 338 (2011) (stating that the presence of counsel is required for all felony prosecutions and misdemeanor prosecutions that result in incarceration).

9. See Virginia E. Sloan et al., *Gideon’s Unfulfilled Mandate: Time for a New Consensus*, HUM. RTS., Winter 2004, at 3, 3 (explaining that *Gideon* does not require quality standards for public defenders).

10. See Cara H. Drinan, *The Third Generation of Indigent Defense Litigation*, 33 N.Y.U. REV. L. & SOC. CHANGE 427, 428 (2009); Stephen F. Hanlon, *State Constitutional Challenges to Indigent Defense Systems*, 75 MO. L. REV. 751, 754–56 (2010); Wayne A. Logan, *Litigating the Ghost of Gideon in Florida: Separation of Powers as a Tool to Achieve Indigent Defense Reform*, 75 MO. L. REV. 885, 885 (2010); Ronald F. Wright, *Parity of Resources for Defense Counsel and the Reach of Public Choice Theory*, 90 IOWA L. REV. 219, 221–22 (2004).

11. *Strickland v. Washington*, 466 U.S. 668 (1984); see also *Missouri v. Frye*, 132 S. Ct. 1399 (2012); *Lafler v. Cooper*, 132 S. Ct. 1376 (2012).

inadequate performance prejudiced the client.¹² But for reasons explored in detail elsewhere, the record about a lawyer's performance can be minimal.¹³ The judge compares one lawyer's activities to the presumed ordinary activities of a comparison pool of lawyers and completely ignores the results obtained for the defendant at bar or for any other clients of the defense attorney in question. The constitutional standard tries to prevent only the weakest performances, as measured by a cursory review of the attorney's amount and type of activity, rather than the results achieved. It does so based on presumptions about activity levels, not data about results. As a method of improving the quality of the median performance for defense attorneys, the constitutional doctrine is a dead end.

Judges sometimes make more nuanced judgments about the quality of defense lawyers when they appoint attorneys to represent indigent defendants.¹⁴ A judge might perform ad hoc evaluations of local attorneys, saving the more skillful lawyers for the most serious cases.¹⁵ It is also common for local rules of court to reserve serious felony trials for attorneys with second-chair experience in comparable cases.¹⁶ In short, the appointment decision rest on unquantifiable impressions of attorney quality (in ad hoc jurisdictions) or on crude measures of past experience (rules requiring a certain number of prior trials).¹⁷

12. *Strickland*, 466 U.S. at 687.

13. See Laurence A. Benner, *The Presumption of Guilt: Systemic Factors that Contribute to Ineffective Assistance of Counsel in California*, 45 CAL. W. L. REV. 263, 322 (2009) (explaining problems with demonstrating prejudice to the defendant when facts establishing innocence are left off the record due to attorney incompetence); Eve Brensike Primus, *The Illusory Right to Counsel*, 37 OHIO N.U. L. REV. 596, 606–07 (2011) (noting the difficulty of demonstrating ineffective assistance of counsel when few errors actually occur on the record).

14. See Bernhard, *supra* note 4, at 846–47 (describing qualification standards for appointed counsel).

15. See *id.* at 847 (“The more challenging or severe the class of crime, the more extensive experience the applicant must have in order to be assigned to the category of case.”).

16. See MARC L. MILLER & RONALD F. WRIGHT, *CRIMINAL PROCEDURES: CASES, STATUTES, AND EXECUTIVE MATERIALS* 812 (4th ed. 2011) (reprinting the appointment rule from Cuyahoga County, Ohio).

17. Cf. Bernhard, *supra* note 4, at 847 (describing evaluation criteria based on prior experience).

These judgments, unlike the assessments that judges make under *Gideon* and *Strickland*, could improve if the judge had richer information available about the performance of individual attorneys. The judge guesses about the proper attorney to appoint based on such thin evidence because the evidence is expensive to develop, not because it is irrelevant.

Just as the judge makes broad-brush evaluations of defense attorneys, prosecutors keep an eye on the quality of the defense lawyers they face. In rare cases, the arrival of a high-quality defense attorney in the case might convince the prosecutor to assign more office resources to that case.¹⁸ At the other extreme, prosecutors remain alert to cases in which an exceptionally weak defense lawyer represents the defendant.¹⁹ Although there is some variety of practice on this matter, at least some prosecutors feel bound to watch for grossly inadequate counsel and to take steps to protect the defendant's rights in extreme cases if the defense lawyer does not perform the job.²⁰ Again, the prosecutor's evaluation of the defense counsel looks to the amount and type of attorney activity, not results. The prosecutor's assessment only becomes relevant in the most extremely inadequate performances.

B. Organizational and Client Guesswork

The more thorough evaluations of defense lawyers come from within their own organizations, whether it be public defender offices or private law offices. The culture within a defense organization is ultimately more important than constitutional

18. See generally Vanessa Merton, *What Do You Do When You Meet a "Walking Violation of the Sixth Amendment" if You're Trying to Put That Lawyer's Client in Jail?*, 69 FORDHAM L. REV. 997 (2000).

19. See *id.* at 1039–44 (discussing the propriety of a prosecutor evaluating and commenting on the poor quality of defense counsel).

20. See, e.g., N.C. RULES OF PROF'L CONDUCT R. 3.8 cmt. 1 (2006) ("A prosecutor has the responsibility of a minister of justice This responsibility carries with it specific obligations to see that the defendant is accorded procedural justice and that guilt is decided upon the basis of sufficient evidence. Precisely how far the prosecutor is required to go in this direction is a matter of debate and varies in different jurisdictions.").

doctrine or legal standards as a driver of quality legal representation.²¹

The most important employers of criminal defense lawyers are public defender organizations, which are spreading to cover larger proportions of all criminal defendants.²² The organizations themselves are becoming larger and more bureaucratic, as individual offices grow larger and statewide supervision and funding builds network links among formerly separate local offices.²³ This organizational growth both increases the capacity to collect information about attorneys and increases the value of analyzing data to improve overall performance.²⁴

In a world of thin resources for criminal defense—a world that *Gideon* institutionalized—managers in these offices need to evaluate the quality of their attorneys.²⁵ If they identify weaker attorneys, the managers can offer the training and support needed to shore up trouble areas. An ability to spot the most talented and effective attorneys is important in choosing mid-level leadership positions in the office, encouraging mentor relationships, and assigning the high-stakes cases that arrive in the office.

While individual attorney evaluations are important within public defender organizations, the current basis for these evaluations is anecdotal and intuitive. Individual managers might receive episodic feedback from judges, clients, or even prosecutors about the quality of an attorney's work. The

21. See Jonathan Rapping, *Directing the Winds of Change: Using Organizational Culture to Reform Indigent Defense*, 9 LOY. J. PUB. INT. L. 177, 181 (2009) (stressing the importance of organizational culture).

22. DONALD J. FAROLE, JR. & LYNN LANGTON, COUNTY-BASED AND LOCAL PUBLIC DEFENDER OFFICES, 2007, at 4 (2010) (describing caseload of county public defender offices); CAROLINE WOLF HARLOW, BUREAU OF JUSTICE STATISTICS, DEFENSE COUNSEL IN CRIMINAL CASES 2 (2000) (comparing the number of federal criminal cases handled by panel attorneys and Federal Defender Organizations); LYNN LANGTON & DONALD J. FAROLE, JR., STATE PUBLIC DEFENDER PROGRAMS, 2007, at 3 (2010).

23. See Ronald F. Wright, Padilla *and the Delivery of Integrated Criminal Defense*, 58 UCLA L. REV. 1515, 1524–27 (2011).

24. Cf. MAX WEBER, ECONOMY AND SOCIETY 973 (1968) (listing reasons for the superiority of bureaucratic organization over other types of organizations).

25. See Bernhard, *supra* note 4, at 840–43 (noting the need for improved training, supervision, and evaluation of public defenders).

supervisor might observe a small sample of the attorney's work, perhaps by sitting in on a trial or by auditing files in particularly important cases.

Managers in public defense organizations do evaluate individual attorneys, but they typically do not rely on quantitative measures.²⁶ When it comes to quantitative measures, these managers more often use data to evaluate defender offices as a whole, rather than relying on statistics to inform their assessment of particular staff attorneys.²⁷ For instance, a few public defender services collect agency-wide statistics that track levels of attorney activity and early contact with clients, such as frequency of visitation with clients during the earliest phases of the representation.²⁸ Organizations also find real value in weighing the caseloads assigned to each attorney and clocking the typical speed of case disposition. They use these metrics to build the political case for continued or increased funding.²⁹

The need for data-driven evaluation is less obvious for criminal defense attorneys who work for private law firms. A disproportionate number of criminal litigators work in solo practice or in small firms.³⁰ They do not adopt the heavy

26. For a description of an evaluation process for individual attorneys in a defender organization, see JOHN S. DIGIACINTO, ANNUAL REPORT FISCAL YEAR 2010–2011 TO THE BOARD OF SUPERVISORS SAN MATEO COUNTY 39–44 (2011); see also *Standards of Practice for Indigent Criminal Defense*, VA. INDIGENT DEF. COMMISSION, <http://www.indigentdefense.virginia.gov/StandofPracEnf.htm> (last visited Apr. 2, 2013) (listing standards of practice for public defenders) (on file with the Washington and Lee Law Review).

27. See FAROLE & LANGTON, *supra* note 22, at 3–6 (using statistics to evaluate public defender offices as a whole, rather than on the basis of individual attorneys).

28. See Erica J. Hashimoto, *Assessing the Indigent Defense System* 12–16 (Sept. 2010), <http://www.acslaw.org/files/Hashimoto%20Indigent%20Defense.pdf> (providing statistics on the amount of contact between counsel and clients).

29. See Heidi Reamer Anderson, *Qualitative Assessments of Effective Assistance of Counsel*, 51 WASHBURN L.J. 571, 577–80 (2012) (describing public defenders' use of quantitative measurements); Wright, *supra* note 10, at 236–41; Spangenberg Grp., *Keeping Defender Workloads Manageable*, <https://www.ncjrs.gov/pdffiles1/bja/185632.pdf> (listing standards for excessive caseload).

30. Cf. *Criminal Litigation Careers: The Private Defense Attorney*, VAULT, <http://www.vault.com/articles/Criminal-Litigation-Careers-The-Private-Defense-Attorney-22541018.html> (last visited Apr. 2, 2013) (“Most criminal defense firms

monitoring practices that larger law firms use to track hours and the quality of work.³¹ Nor does there appear to be much demand among the leaders of smaller firms for better tools to monitor and evaluate the defense attorneys within the firm.³²

On the other hand, it is easy to imagine that clients would have a sharp interest in evaluating and comparing defense attorneys. And it seems clear that results in past cases would matter greatly to these potential future clients. Perhaps more than any other group, clients would benefit from some meaningful way to evaluate the quality of defense attorneys potentially available to them. Meanwhile, clients have the least access to meaningful information that would allow them to compare one attorney to another. Perhaps the client receives a recommendation from a friend, often based on a favorable experience of one previous client. Or perhaps the client draws conclusions about the attorney's quality based on truly irrelevant clues. Whatever the source, clients enter the attorney-client relationship with information too feeble to support any weight at all.

III. Evaluation in a Data-Impoverished Context

If the need for better measures of attorney quality is clear enough, the way to get better measures is not so clear. In fact, there are barriers to meaningful, results-based evaluations of attorneys generally.³³ Those challenges are especially pronounced for criminal defense attorneys.³⁴

are smaller than their corporate counterparts.”) (on file with the Washington and Lee Law Review).

31. See W. Keith Shannon & Russell J. White, *TQM—Ready or Not*, S.C. LAW., Jan.–Feb. 1994, at 11, 15 (“Many lawyers in solo and small firm practices tend to dismiss the idea of [quality analysis] as being designed for use only in larger firms.”).

32. *Id.*

33. See William H. Simon, *Where Is the “Quality Movement” in Law Practice?*, 2012 WIS. L. REV. 387, 402–03 (describing how professional culture makes attorneys reluctant to be evaluated based on performance).

34. See Kim Taylor-Thompson, *Tuning Up Gideon’s Trumpet*, 71 FORDHAM L. REV. 1461, 1487–88 (2003) (discussing a lack of information exchange between defense attorneys and clients).

Lawyers provide a complex service, offering a blend of legal and practical advice to clients who present an infinite variety of problems.³⁵ The bottom line of a “win” or a “loss” is difficult to define, and does not begin to tell the full story of how an attorney could add value for a given client.³⁶

In many jurisdictions, the ethics rules embody this reluctance to measure attorney quality by results. For instance, the rules related to advertising for clients build on the assumption that any “guarantee” of results would be misleading.³⁷ Client testimonials about favorable outcomes in past cases are also suspect and must include a disclaimer that past outcomes do not guarantee future results.³⁸ Even direct comparisons of one lawyer’s skills to another’s raise concerns for bar enforcers, who insist that such advertisements base their comparisons on verifiable facts.³⁹

While the complexity of legal services presents a challenge to rating lawyers in many subject areas, an additional problem appears in the criminal context. Criminal courts operate with extremely uneven data. The courts handle large numbers of defendants with underfunded and antiquated computers.⁴⁰ The people who spend the most time in criminal courts, on behalf of themselves and their family members charged with crimes, are not rich, well-organized, or politically influential. The funding for criminal court data is anemic, and the political landscape keeps it that way.

Given this data environment, it would be especially challenging to learn in detail about a criminal defense lawyer’s work based on court or office records. Much of the recorded

35. See Wright, *supra* note 23, at 1516–17 (discussing how civil issues now often accompany criminal issues).

36. Cf. Simon, *supra* note 33, at 402–04 (describing resistance to the evaluation of lawyer quality).

37. See, e.g., MODEL RULES OF PROF’L CONDUCT R. 7.1 cmt. 3 (2012).

38. See *id.*; N.C. RULES OF PROF’L CONDUCT R. 7.1 cmt. 3 (2003).

39. See *supra* note 38 and accompanying text.

40. See William T. Robinson III, *Criminal Justice Reforms Enhance Public Safety and Strengthen Our Courts*, NAT’L CTR. FOR ST. CTS., <http://www.ncsc.org/sitecore/content/microsites/future-trends-2012/home/Leadership-and-the-Courts/4-1-Enhancing-Public-Safety.aspx> (last visited Apr. 2, 2013) (discussing the negative impact of budget cuts for courts) (on file with the Washington and Lee Law Review).

information is only captured on paper, with surprisingly little of it transferred to electronic format.⁴¹ The information that does make it into electronic format lands in separate data systems that do not communicate with each other.⁴² The quality of data available for any given criminal case is shallow and appalling, even though the sheer number of cases is impressive.

Talent evaluators in other contexts have dealt with this problem, finding ways to draw restrained but meaningful insights from the limited data. As dramatized in *Moneyball*, a manager in professional sports can use performance data to evaluate talent and to develop players.⁴³ For instance, hitters with strong on-base percentages and strike zone discipline caught the attention of the Oakland A's.⁴⁴ Minor league players in the A's organization learned at every level the importance of walks and forcing the opposing pitcher to throw extra pitches.⁴⁵ Statistical analysis has also taken a heavier role in player selection and game planning in other sports, such as basketball—sports that traditionally placed less emphasis on collecting or analyzing performance statistics.⁴⁶

Statistical ratings have also become salient (and controversial) for other complex service industries, such as education and health care.⁴⁷ School systems show a growing

41. Melissa Maynard, *State Courts Replacing Paper with Electronic Record Keeping*, PEW CTR. ON STS. (Apr. 24, 2012), <http://www.pewstates.org/projects/stateline/headlines/state-courts-replacing-paper-with-electronic-record-keeping-85899382276> (last visited Apr. 2, 2013) (explaining the difficulty in switching to electronic record keeping) (on file with the Washington and Lee Law Review).

42. *See id.* (discussing the lack of electronic records preservation programs in Texas).

43. *Moneyball*, *supra* note 1.

44. *Id.*

45. *Id.*

46. *See* Brian Mossop, *Basketball Isn't a Sport. It's a Statistical Network*, WIRED (Dec. 7, 2012, 2:45 PM), <http://www.wired.com/playbook/2012/12/basketball-network-analysis/> (last visited Apr. 2, 2013) (discussing the use of statistical analysis in basketball) (on file with the Washington and Lee Law Review).

47. *See* Deborah Kenny, *Want to Ruin Teaching? Give Ratings*, N.Y. TIMES (Oct. 14, 2012), <http://www.nytimes.com/2012/10/15/opinion/want-to-ruin-teaching-give-ratings.html> (last visited Apr. 2, 2013) (arguing against government evaluation of teachers) (on file with the Washington and Lee Law Review); Eric Zorn, *Why Teachers Have Test Anxiety*, CHI. TRIB. (Sept. 14, 2012),

interest in quantitative performance evaluations of teachers in primary and secondary schools.⁴⁸ Those performance evaluations, once based on ad hoc observations by school principals, increasingly include a component based on the test score performance of the teacher's students.⁴⁹ There is dispute about an appropriate weight to place on test scores, as compared to personal observation of teaching and evaluation of lesson plans and other documentation of the teacher's work.⁵⁰ The advocates for ratings, such as student test performance, concede that they must control for socioeconomic status and other predictors of student performance to make realistic judgments about how much "value" the teacher adds above and beyond the improvements in scores that one might expect from a given group of students.⁵¹ Teacher ratings that include test score performance also create dilemmas about who should have access to those ratings.⁵² It is a complicated landscape for teacher evaluation, but ratings based in part on student test performance appear to be a settled part of the scene.⁵³

Data-based evaluations have also arrived in health care. Hospitals keep detailed records of treatment decisions and

http://articles.chicagotribune.com/2012-09-14/news/ct-oped-0914-zorn-20120914_1_value-added-approach-bad-teachers-teacher-evaluation-process (last visited Apr. 2, 2013) (expressing skepticism at the use of test scores in teacher evaluations) (on file with the Washington and Lee Law Review); *Can School Performance Be Measured Fairly*, in *Room for Debate*, N.Y. TIMES (July 29, 2012), <http://www.nytimes.com/roomfordebate/2012/07/29/can-school-performance-be-measured-fairly/> (last visited Apr. 2, 2013) (describing the debate over evaluation of school performance) (on file with the Washington and Lee Law Review).

48. See Zorn, *supra* note 47 (explaining a public school system's wish to use "value added" analysis to evaluate teachers).

49. See *id.* (discussing the role of test scores in evaluations).

50. See Kenny, *supra* note 47 (raising objections to the use of test scores).

51. Zorn, *supra* note 47.

52. Cf. Adam Shanks, *Parents Will Have Access to Teacher Evaluations*, LEGIS. GAZETTE (June 25, 2012), <http://www.legislativegazette.com/Articles-Top-Stories-c-2012-06-25-81858.113122-Parents-will-have-access-to-teacher-evaluations.html> (last visited Apr. 2, 2013) (describing debate over parents' access to teacher evaluations) (on file with the Washington and Lee Law Review).

53. See Zorn, *supra* note 47 (noting that performance-based evaluations will soon be used in every state).

patient outcomes.⁵⁴ Some of those data are only now taking electronic form, and many of the electronic data are incompatible and inaccessible (the same situation we find in state criminal justice).⁵⁵ Nevertheless, this is a changing data landscape. As more data about treatment and patient outcomes become available, it becomes possible to rate the performance of particular sectors of the industry, or particular hospitals, or even particular doctors.⁵⁶ Particularly, given the recently expanded access to the enormous database showing treatments and outcomes for Medicare and Medicaid patients, analysts have begun to rate particular practice areas and hospitals and to release those ratings to the public.⁵⁷ The hospitals often take issue with the comparison groups that the researchers choose or other research design questions.⁵⁸ Still, the use of performance data as one component in evaluations of a complex service appears to have staying power in the health care field.

54. See *Improved Diagnostics and Patient Outcomes*, HEALTHIT, <http://www.healthit.gov/providers-professionals/improved-diagnostics-patient-outcomes> (last visited Apr. 2, 2013) (describing benefits of electronic health records) (on file with the Washington and Lee Law Review).

55. See Fig Gungor, *Disadvantages of Electronic Medical Records*, ONESOURCE (Nov. 11, 2011), <http://www.onesourcedoc.com/blog/bid/71535/Disadvantages-of-Electronic-Medical-Records> (last visited Apr. 2, 2013) (explaining how disparate systems are difficult to synchronize and result in duplications) (on file with the Washington and Lee Law Review).

56. See *About Leapfrog*, LEAPFROG GROUP, http://www.leapfroggroup.org/about_leapfrog (last visited Apr. 2, 2013) (discussing performance evaluation for health care) (on file with the Washington and Lee Law Review).

57. See Jordan Rau, *Patient Ratings to Affect Medicare Payments to Hospitals*, WASH. POST (Apr. 28, 2011), http://articles.washingtonpost.com/2011-04-28/national/35262055_1_hospital-room-small-community-hospitals-hospitals-conduct (last visited Apr. 2, 2013) (“Medicare has been publishing patient-satisfaction scores on its Hospital Compare Web site since 2008 . . .”) (on file with the Washington and Lee Law Review).

58. Jordan Rau, *Teaching Hospitals Decry Medicare’s Low Safety Rankings*, BOSTON.COM (Feb. 23, 2013), http://www.boston.com/lifestyle/health/articles/2012/02/13/teaching_hospitals_challenge_low_safety_ranking_by_medicare/ (last visited Apr. 2, 2013) (describing the negative reaction to hospital evaluations) (on file with the Washington and Lee Law Review); *National Rankings for Hospitals*, HEALTH INSIGHT, <http://www.healthinsight.org/Internal/HospitalPerformanceRankings.html> (last visited Apr. 2, 2013) (explaining quality measures) (on file with the Washington and Lee Law Review).

The people who evaluate complex services such as education or health care find it worthwhile to gather and analyze performance data. They do so despite the limitations of the data and the need to combine the statistical ratings with more in-depth personal observation. Could the same hold true for the work of criminal defense attorneys?

IV. A Bottom-Line System to Rate Attorneys

In this Part, we describe past efforts to use performance data in evaluating criminal defense lawyers. Building on that history, we describe our design for a system to rate North Carolina's criminal defense lawyers, based on publicly available data.

A. Comparing Public Defenders to Private Defense Attorneys

The use of data to rate the effectiveness of defense attorneys has long fascinated legal scholars. Dozens of evaluations have appeared in print over the years, going back at least to the 1930s.⁵⁹ Most of those efforts compare three sets of attorneys, evaluated as groups: public defenders, appointed counsel from private practice, and retained attorneys. These studies tie into perennial legislative policy debates about whether to create public defender services, and how to get the best results from limited public funds.⁶⁰

Three design features of the quantitative studies interest us here. The authors conducting the study must choose a measure of attorney success; to use the lingo of social science empiricism, they must decide how to operationalize the dependent variable. In addition, the study design must select which aspects of the attorneys' background and environment might contribute to successful outcomes; these are the independent variables of

59. Floyd Feeney & Patrick Johnson, *Public Defenders, Assigned Counsel, Retained Counsel: Does the Type of Criminal Defense Matter?*, 22 RUTGERS L.J. 361, 416 tbl.3 (1991) (listing studies dating back to 1935).

60. For an excellent review of the enormous number of early studies, see *id.* at 365–78.

interest. Finally, the researchers must decide which influences, apart from attorney effort, might contribute in some degree to the chosen output; that is, the studies must include independent variables that interact with the variables of interest.

As for the first design feature, the earliest studies used a variety of measures for attorney quality. Some focused on the amount and type of activity of an attorney, such as the number of motions filed in a case, the number of cases that the lawyer tried, or various measures of speedy processing of cases.⁶¹ Others looked to outcomes—the proportion of defendants convicted—with attorneys rated more highly when they achieved lower conviction rates.⁶² The early studies also inquired about the type of sentence that the defendant received, giving attorneys credit for those defendants who avoided prison sentences.⁶³

61. See ROGER A. HANSON ET AL., *INDIGENT DEFENDERS: GET THE JOB DONE AND DONE WELL* 39–48 (1992) (total time of disposition); STEVEN K. SMITH & CAROL J. DEFRANCES, U.S. DEPT. OF JUSTICE, *INDIGENT DEFENSE* 4 tbl.7 (1996) (measure based on lawyer meeting client within one week of arrest); *id.* at 45 tbl.11 (trial rates); Michael McConville & Chester L. Mirsky, *Criminal Defense of the Poor in New York City*, 15 N.Y.U. REV. L. & SOC. CHANGE 581, 766–70 (1986–1987) (number of motions); see generally Note, *Representation of Indigents in California—A Field Study of the Public Defender and Assigned Counsel Systems*, 13 STAN. L. REV. 522 (1960).

62. See Joyce S. Sterling, *Retained Counsel Versus the Public Defender: The Impact of Type of Counsel on Charge Bargaining*, in *THE DEFENSE COUNSEL* 151, 160–62 (William McDonald ed., 1983) (comparing outcomes between types of attorneys); Dean J. Champion, *Private Counsels and Public Defenders: A Look at Weak Cases, Prior Records and Leniency in Bargaining*, 17 J. CRIM. JUST. 253, 258 tbl.1 (1989) (comparing plea bargain outcomes between private attorneys and public defenders).

63. See Richard L. Grier, *Analysis and Comparison of the Assigned Counsel and Public Defender Systems*, 49 N.C. L. REV. 705, 714 (1971) (comparing the percentage of convictions given probation or suspended sentence for public defenders and assigned counsel); Pauline Houlden & Steven Balkin, *Quality and Cost Comparisons of Private Bar Indigent Defense Systems: Contract vs. Ordered Assigned Counsel*, 76 J. CRIM. L. & CRIMINOLOGY 176, 179 (1985) (comparing convictions and sentences for clients of assigned counsel and public defenders); Stuart S. Nagel, *Effects of Alternative Types of Counsel on Criminal Procedure Treatment*, 48 IND. L.J. 404, 414–18 (1973) (comparing results for clients with hired counsel and provided counsel); Jean G. Taylor et al., *An Analysis of Defense Counsel in the Processing of Felony Defendants in Denver, Colorado*, 50 DENV. U. L. REV. 9, 19–21 (1973) (same); Jennifer Bennett Shinall, Note, *Slipping Away from Justice: The Effect of Attorney Skill on Trial Outcomes*, 63 VAND. L. REV. 267, 270–78 (2010) (describing prior research).

More recent studies homed in on the length of prison or jail terms and other measures of sentence *severity* as the measure of attorney quality.⁶⁴ This measure of attorney performance has several advantages. First, it asks a question that court data can answer: what sentence (if any) did the court impose on the defendant at the end of the trial court proceedings? Second, this measure asks something that is on the mind of many defendants. While the bottom line of punishment is not the only concern of defendants, clients ultimately want to know what penalty will result, not how many motions the attorney will file.⁶⁵ If two clients are convicted of the same crime, despite the efforts of equally active and empathetic defense lawyers, the client who receives the lower sentence would presumably rate that attorney more highly.⁶⁶

After choosing a type of activity or outcome that approximates an overall high quality of lawyering, the researcher also must estimate how much of that “success” to attribute to the lawyer, and how much to attribute to other factors, such as the seriousness of the charge, the offender’s prior record, other offender characteristics, the practices of the prosecutor or the judge, and so forth.⁶⁷ Put another way, a researcher who hopes to rate the quality of defense counsel must set expectations. How

64. See James M. Anderson & Paul Heaton, *How Much Difference Does the Lawyer Make? The Effect of Defense Counsel on Murder Case Outcomes*, 122 YALE L.J. 154, 178–87 (2012) (using multiple measures of attorney success, including conviction rate, life sentence versus death penalty, and length of prison term); Morris B. Hoffman, Paul H. Rubin & Joanna M. Shepherd, *An Empirical Study of Public Defender Effectiveness: Self-Selection by the “Marginally Indigent,”* 3 OHIO ST. J. CRIM. L. 223, 231 (2005) (studying different outcomes in plea bargaining). *But see* Richard Posner & Albert Yoon, *What Judges Think of the Quality of Legal Representation*, 63 STAN. L. REV. 317, 324–39 (2011) (using quality measures other than sentencing outcomes, such as findings based on judicial surveys).

65. See Hoffman et al., *supra* note 64, at 225 (“What criminal defendants care most about . . . is the actual outcome of the case . . .”).

66. Interview with Robin Steinberg, Exec. Dir., Bronx Defenders, in Lexington, Va. (Nov. 8–9, 2012); *cf.* TOM R. TYLER, WHY PEOPLE OBEY THE LAW 115–24 (2006) (examining the factors affecting whether someone believes that process is fair).

67. See Anderson & Heaton, *supra* note 64, at 156 (testing the criminal justice system’s sensitivity to the quality of defense counsel).

much success should one expect for certain types of cases, and how much “value added” can one fairly attribute to higher quality defense lawyering, given the other constraints of the case? How does one measure how well the attorney performs with the materials at hand?

Again, studies have evolved over time. The earliest research simply calculated the relevant measure of success for the different types of defense lawyers.⁶⁸ They did not control for factors related to the “raw materials” presented to the defense attorney. Later studies recognized this problem and started to control for the seriousness of the initial charges against the defendant. The prior criminal record of the defendant also became a feature in the studies, along with other defendant characteristics.⁶⁹

In addition to evolving measures of attorney success and increasing efforts to separate the quality of lawyering from the nature of the “raw material” presented to the attorney, researchers over the years began to ask about the personal backgrounds and work environments of defense attorneys that might explain the quality of lawyering they could offer.⁷⁰ For decades, the only input that attracted any attention was the form of organization: public defender offices versus appointed private counsel versus retained private counsel.⁷¹ Some of the more

68. See Feeney & Johnson, *supra* note 59, at 365–66.

69. See MICHAEL ROACH, EXPLAINING THE OUTCOME GAP BETWEEN DIFFERENT TYPES OF INDIGENT DEFENSE COUNSEL: ADVERSE SELECTION AND MORAL HAZARD EFFECTS 10–11 (2011) (including defendant characteristics in his study).

70. See Radha Iyengar, *An Analysis of the Performance of Federal Indigent Defense Counsel* 4 (Nat'l Bureau of Econ. Research, Working Paper No. 13187, 2007) (comparing attorney backgrounds).

71. That environmental factor remains the independent variable of interest in some recent studies. See THOMAS H. COHEN, WHO'S BETTER AT DEFENDING CRIMINALS? DOES TYPE OF DEFENSE ATTORNEY MATTER IN TERMS OF PRODUCING FAVORABLE CASE OUTCOMES 2 (2011) (comparing public defenders, assigned counsel, and private attorneys); David Allan Felice, *Justice Rationed: A Look at Alabama's Present Indigent Defense System with a Vision Towards Change*, 52 ALA. L. REV. 975, 991–95 (2000) (same); Kuo-Chang Huan, Kong-Pin Chen & Chang-Ching Lin, *Does the Type of Criminal Defense Counsel Affect Case Outcomes? A Natural Experiment in Taiwan*, 30 INT'L REV. L. & ECON. 113, 117–18 (2010) (comparing public defenders and appointed counsel); Peter A. Joy & Kevin C. McMunigal, *Does the Lawyer Make a Difference? Public Defender Versus*

recent studies, however, also ask about factors such as the amount of experience and the demographics of individual attorneys as possible explanations for stronger lawyer performances.⁷²

The quantitative studies of lawyer effectiveness over the years have produced varied results. Some indicate that public defenders are more effective than retained counsel, while others conclude the opposite.⁷³ Some indicate that more experienced attorneys tend to be more effective, while others find no such effect.⁷⁴ Some find evidence that lawyers with particular backgrounds get better results for criminal defendants; for instance, Abrams and Yoon conclude that Hispanic attorneys in the Las Vegas public defender office got better results for criminal defendants, possibly because they were not hampered by language barriers with the clients and witnesses.⁷⁵

Such conflicting results in the historical studies do not surprise us. Given the amazing variety of models that different jurisdictions use to deliver criminal defense services around the country, it is easy to believe that different factors explain the success of attorneys in those different systems. The major changes over time in the structure of criminal defense organizations also suggest that the formula for success changes over the years.

The task of a manager of criminal defense services in a particular place and a given time is not to declare the universal conditions necessary to do the job well. The real challenge is to measure local conditions and to make the most of the resources at hand. A clear-eyed view of the available data can contribute to that wise management.

Appointed Counsel, A.B.A. CRIM. JUST. (Spring 2012), http://www.americanbar.org/content/dam/aba/publications/criminal_justice_magazine/sp12_ethics.auth_checkdam.pdf (same).

72. See Iyengar, *supra* note 70, at 4–6 (analyzing attorney characteristics).

73. See Anderson & Heaton, *supra* note 64, at 157 n.6.

74. David S. Abrams & Albert H. Yoon, *The Luck of the Draw: Using Random Case Assignment to Investigate Attorney Ability*, 74 U. CHI. L. REV. 1145, 1145 (2007).

75. *Id.* at 1175.

B. Attorney Ratings in a Guideline Sentencing World

The rich history of quantitative studies of defense counsel quality is a hopeful sign for our demonstration project; it suggests that such an exercise can bear fruit. At the same time, the challenges in building and designing a credible quantitative study are daunting. A database that is assembled one time only, based on the full-time effort of researchers putting together data sources that normally do not integrate, would be difficult to replicate for managers in hundreds of offices using real-world budgets, schedules, and expertise. In some cases, the researchers must extract case-level information from paper files to include in the data. What works under social science laboratory conditions may not work in the field.

We believe, nevertheless, that we can demonstrate the practicality of performance-based evaluations under field conditions. Two factors in our test jurisdiction—features that are not especially unusual—work in our favor. First, North Carolina collects some statewide data about criminal court operations in ACIS, the Automated Criminal Infractions System.⁷⁶ Granted, ACIS is a balky older program with surprising gaps (the core of the program was written in 1968), but it does offer to actors in the system an electronic record of charges filed, disposition of those charges, sentences imposed, and the names of key actors in each case. It also allows us to compare, when appropriate, the performance of defense lawyers across different districts in the state, giving us a better vantage point to notice the differences that office organization and other environmental distinctions might make for attorney performance.

The presence of sentencing guidelines in North Carolina offers a second natural design advantage.⁷⁷ Sentencing guidelines

76. See *Courtflow*, N.C. CTS. SYS., <http://www.nccourts.org/Courts/CRS/Technology/Applications/CourtFlow.asp> (last visited Apr. 2, 2013) (describing ACIS) (on file with the Washington and Lee Law Review).

77. See N.C. GEN. STAT. ANN. § 15A-1340.14 (West 2012) (establishing sentencing guidelines according to prior criminal history). Almost half of the states operate some form of sentencing guidelines. See Richard S. Frase, *State Sentencing Guidelines: Diversity, Consensus, and Unresolved Policy Issues*, 105 COLUM. L. REV. 1190, 1191 (2005) (“Sentencing guidelines have been adopted in at least eighteen states and the District of Columbia . . .”). For some speculation about

encourage system actors to use common sentencing practices or to declare explicitly why the case at hand deserves special treatment.⁷⁸ Guidelines therefore make it easier to predict the expected sentencing effects of factors such as the crime of conviction or the defendant's prior criminal record. In short, sentencing guidelines allow researchers to measure more precisely some features of the raw materials that attorneys inherit.⁷⁹

What follows, then, is a step-by-step account of how we converted state court data into a crude but serviceable measure of some relevant aspects of criminal defense work.

1. Choice of Attorney Quality Measure

Like other recent researchers, we favor a quality metric focused on the sentencing reductions that the defense lawyer obtains. Such a measure focuses on an outcome that matters to defendants, and one that captures the negotiation skills that are central to modern criminal practice.

Such a one-dimensional measure would not be appropriate when evaluating the work of prosecutors.⁸⁰ The prosecutor's work necessarily balances competing interests of immediate crime victims, potential future victims, law enforcement agencies, the public's interest in a sense of public safety, and even the defendant's long-term interests. Prosecutors, who have invested serious effort in multi-factor performance measures, recognize

some potential advantages of measuring defense counsel performance in a sentencing guideline jurisdiction, see Douglas A. Berman, *From Lawlessness to Too Much Law? Exploring the Risk of Disparity from Differences in Defense Counsel Under Guidelines Sentencing*, 87 IOWA L. REV. 435, 439 (2002).

78. See *id.* § 15A-1340.16 (discussing aggravating and mitigating factors).

79. Abrams & Yoon, *supra* note 74, at 1149, addressed this selection bias issue in the caseload by studying a public defender office that assigns cases to attorneys on a random basis, a relatively unusual situation. See also Anderson & Heaton, *supra* note 64, at 159 (studying a public defender office that randomly assigns murder cases).

80. See Mary De Ming Fan, *Disciplining Criminal Justice: The Peril Amid the Promise of Numbers*, 26 YALE L. & POL'Y REV. 1, 57–65 (2007) (addressing problems with evaluation based solely on convictions).

this complexity.⁸¹ The severity of sentences that a prosecutor obtains, standing alone, would be an unacceptable measure of quality.

This difficulty is less acute, however, for defense attorneys. The defendant's lawyer pursues the interests of the client. That duty raises subtle questions about the full range of legal problems (civil and criminal) that might face a defendant, and the lawyer may want to help the client consider short-term versus long-term interests.⁸² But the sentencing outcome for a given criminal case, under mainstream views of the defense attorney's job, is central to the attorney's success. Sometimes the attorney holds a weak hand, but it is reasonably clear what the defense lawyer should try to accomplish with the cards as dealt.

2. *Starting Point and Value Added*

As we calculate the sentence reductions that the attorney achieves in the case, we must set both the expected starting point and the end point for the sentence. In North Carolina, ACIS offers the arraignment charge against a defendant as a starting point.⁸³ We can estimate the value of the case before the defense attorney ever gets involved.

Sentencing guidelines allow us to make a reasonable estimate for the sentence that would ordinarily attach to any combination of a particular charge with a particular criminal history category (a factor also captured in ACIS). Knowing the offense class of the arraignment charge and the defendant's criminal history category, we calculate the "starting point"

81. See M. Elaine Nugent-Borokove et al., *Performance Measures for Prosecutors: Findings from the Application of Performance Measures in Two Prosecutors' Offices*, NAT'L DISTRICT ATT'YS ASS'N 1 (Apr. 2007), http://www.ndaa.org/pdf/performance_measures_findings_07.pdf (acknowledging the diverse roles played by prosecutors).

82. See Wright, *supra* note 23, at 1516–17 (noting the inclusion of civil penalties in criminal cases).

83. Our data captures any judicial action taken as of June 1, 2010, in any case originally filed as a felony during calendar year 2006. We chose 2006 in an effort to give most cases a chance to complete their path through the system to sentencing.

sentence as the number of days of incarceration that a judge would impose after selecting the bottom of the “presumptive” range in the relevant box of the sentencing grid (a box that already reflects the seriousness of the offense and the extent of the prior criminal record). We chose this point in the available range because it is the most common choice among North Carolina sentencing judges when exercising their available discretion within this system.⁸⁴ For instance, we assign to breaking and entering (a Class H felony) an expected sentence ranging from five months for a defendant with no criminal history (Category I) up to sixteen months for a defendant with an extensive criminal history (Category VI).

As for the end-point sentence, the ACIS system records the crime of conviction and the type of sentence actually imposed (incarceration or non-incarceration), along with the length of any prison or jail term.⁸⁵ By subtracting the duration of the sentence imposed from the duration of the expected sentence (based on the arraignment charge), we can see the “value added” for a defendant in any given case. The larger the value added, the more successful the defense attorney.

84. See N.C. SENT'G & POL'Y ADVISORY COMM'N, STRUCTURED SENTENCING STATISTICAL REPORT FOR FELONIES AND MISDEMEANORS 18–20 (2012). Another important aspect of the sentencing judge's discretion under state law is the choice between concurrent and consecutive sentences in multiple-count cases. We included in our ratings any sentences listed separately in the ACIS system; a more refined system would screen out the multi-count sentences to be served concurrently.

85. North Carolina sentencing law gives the judge, within some grid boxes, a choice of “community” sentence for correctional programs that do not restrict the offender's physical liberty, or an “intermediate” sentence under correctional programs that restrict the offender's movement less than an active prison term but more than a community sanction would. See N.C. GEN. STAT. ANN. § 15A-1340.11 (West 2012) (defining “community” and “intermediate” sentences). We treated community sentences as a sentence of zero days, and treated intermediate punishments as equivalent in length to an active prison term. In customizing a rating system for a particular jurisdiction, one could alternatively score intermediate and community punishments as some fraction of the equivalent number of days for an active prison sentence.

3. Comparisons Among Cases, Crimes, and Districts

Once a rater knows the amount of sentencing reduction in a given case, how does one judge whether the reduction is a high, low, or normal amount? Our answer is to compare the attorney in a single case with the pool of all attorneys statewide who represented defendants charged with the same offense at arraignment. For each crime in the arraignment charge column, we calculated the statewide average in reductions of sentence in all cases starting out with that charge.⁸⁶ For example, defendants who were arraigned on a charge of breaking and entering received an average reduction of 105 days from their expected starting point sentence to the sentence actually imposed on them. An attorney who achieved reductions greater than 105 days for a defendant initially charged with breaking and entering performed above the statewide average in that case.

Other comparison pools are also possible. Rather than looking to a statewide average within a single crime for attorney performance, one might instead compare one attorney to others working in the same prosecutorial district—those who negotiate against the same prosecutors in the same courthouse environment. The difficulty with this strategy, however, is the problem of small numbers. Any given district might produce too few charges of robbery with a dangerous weapon during a single year to provide a reliable average score for purposes of comparison. Managers who prefer to draw comparisons within their own prosecutorial district might do so by combining data from several years.⁸⁷

86. The average reduction for each criminal history category within each crime would produce a more thorough account of the effect of criminal history, but such calculations would lead to a small numbers problem for most crimes in the system. A single statewide average for all criminal history categories still reflects the amount of movement from an expected starting point that considers criminal history, to an end point that also considers criminal history. The pooling of all the categories does blur any noteworthy differences in the amount of movement that may happen in different criminal history categories.

87. Alternatively, a manager might remain within a single prosecutorial district by using all crimes at a given offense class as the relevant comparison pool. We opted instead to use single crimes on a statewide basis because the opportunities for movement might differ dramatically among crimes within the same offense class. See Ronald F. Wright & Rodney Engen, *The Charging and*

Some types of crimes routinely produced greater movement than other crimes. To make it easier to compare the value added for defendants across different crimes, we calculated the standard deviation for the sentence reductions within each crime as charged at arraignment. Thus, even though the average movement for breaking and entering as an arraignment charge was a reduction of 105 days, while the average movement for robbery with a dangerous weapon was an increase of 563 days, an attorney who performed slightly better than two-thirds of the attorneys in the state on a robbery charge (that is, one standard deviation from the average for that crime) gained the same rating benefit as an attorney who performed slightly better than two-thirds of the attorneys in the state on a breaking and entering charge. This method allows us to weight cases equally, preventing a few high-movement cases from swamping the effects of all the other cases.

4. Compiling the Case Ratings of Individual Attorneys

Our next step is to combine all of the individual scores for an attorney's cases into a single "index" score to estimate the attorney's typical "value added." The manager might calculate the average score for that attorney's cases over a one-year or multi-year period.⁸⁸ A manager who wants to estimate how much value a defense attorney in the office usually adds to a defendant's case could glance down a list of these index scores. The more cases that appear in the database, the more comfortable the manager could be that the ratings reveal something meaningful and are not based on a few atypical cases.

These calculations of sentence movement offer only a partial picture: the manager would want to inquire more carefully about

Sentencing Effects of Depth and Distance in a Criminal Code, 84 N.C. L. REV. 1935, 1970 (2006) ("[W]e see meaningful differences in the magnitude of reductions among different types of crimes, even those originally charged at the same offense class.").

88. If the manager believes that some of the highest and lowest scores in an attorney's portfolio reflect unusual cases that should not carry much weight in the overall rating, the system could rely on the median score rather than the mean.

the reasons behind apparently weak numbers on sentence movement when compared to other attorneys. Given the major differences in cases that the ACIS data does not capture, a manager would be irresponsible to attach any meaning to small differences among attorneys. For larger differences that build up across many cases that an attorney handles, a manager should heed the warning and inquire more closely.

Table 1 illustrates the type of report that a manager of criminal defense attorneys could receive on a routine basis. The scores we report here are based on actual calculations from the ACIS data for the criminal defense attorneys who practice regularly in a single prosecutorial district.⁸⁹ Most of their index scores clump between -1 and 1 because the underlying case scores are expressed in standard deviations. The “N” column indicates the number of sentenced cases appearing in the database for that lawyer; attorney scores based on a higher number of case performances are more likely to reflect real performance differences than an attorney average based on fewer cases.

89. Table 1 includes every attorney in District 18 with more than thirty sentences recorded in our 2006 felony data. Although the scores and the underlying data are genuine, we changed the attorney names in a Dickensian spirit.

Table 1: Sample Report of Attorney Value-Added Scores
 Defense Attorneys in Prosecutorial District 18

ATTORNEY NAME	VALUE ADDED	N	MARGIN OF ERROR, +/-
Agnes Wickfield	0.45	35	0.70
Sydney Carton	0.29	32	0.73
Tommy Traddles	0.28	85	0.16
Amy Dorrit	0.19	31	0.29
Jarvis Lorry	0.18	132	0.13
Thomas Gradgrind	0.15	32	0.22
Betsy Trotwood	0.12	55	0.18
Josiah Tulkinghorn	0.10	69	0.19
Kate Nickleby	0.08	158	0.17
Simon Tappertit	0.03	93	0.08
Esther Summerson	0.02	79	0.16
Wilkins Micawber	0.01	38	0.12
Charity Pecksniff	-0.02	37	0.17
Thomas Lenville	-0.03	55	0.17
Jonas Chuzzlewit	-0.03	64	0.07
Grace Jeddler	-0.03	70	0.13
Mortimer Lightwood	-0.05	90	0.11
Estella Havisham	-0.05	56	0.07
Eugene Wrayburn	-0.07	31	0.27
Elizabeth Hexam	-0.07	57	0.10
Waived	-0.07	191	0.07
Martha Bardell	-0.11	135	0.29
Leicester Dedlock	-0.21	102	0.27
Sampson Brass	-0.23	165	0.27
Uriah Heep	-0.35	34	0.61

The “Margin of Error, +/-” column shows the level of confidence that a manager could place in an attorney’s score in light of the number of cases the attorney handled and the dispersion of individual case scores.⁹⁰ Thus, a manager reading

90. The margin of error for the mean index score for each attorney is calculated at the 95% confidence level. This number is a function of the number

this report could have some confidence that Tommy Traddles performed at a routinely high level during the year in comparison to other attorneys around the state who represented defendants facing the same charges and presenting the same prior criminal records. The report indicates that a manager can have 95% confidence that Traddles's true average score falls somewhere between 0.44 and 0.12. On the other hand, the manager would be more cautious in drawing conclusions about Agnes Wickfield and Sydney Carton because the margin of error for both of those attorneys is larger. In both cases, the small number of cases and the wide dispersion of the scores in that lawyer's portfolio mean that the positive score here might be unreliable, the product of sampling error. In Carton's case, the "true" score probably falls somewhere between 1.02 and -0.44. Given the scores and margins of error reported here, it seems reasonable to assume that Traddles, Dorrit, Lorry, Gradgrind, Trotwood, and Tulkinghorn all turned in positive performances for the year.

On the other hand, a reader of this report might begin to worry about the performance of Leicester Dedlock and Sampson Brass. In both cases, the value added score is a good deal lower than other attorneys who regularly represented defendants in the district, and the margin of error is small enough to put almost the entire confidence interval in the negative range. Although Uriah Heep turned in the weakest average score in the district, the high margin of error should lead a manager to be careful in drawing conclusions about his performance during the year. Note that four attorneys in the district added less value, on average, than the typical defendant who waived defense counsel altogether.⁹¹

The same data in a ratings system could also produce other reports to shed light on the performance of defense counsel in a district. For instance, a report might inform a manager about the performance of defense attorneys in particular crimes, compared

of cases handled and the standard deviation of the scores for that attorney. Larger numbers of cases and smaller standard deviations produce tighter confidence intervals.

91. Cf. Erica J. Hashimoto, *Defending the Right of Self-Representation: An Empirical Look at the Pro Se Felony Defendant*, 85 N.C. L. REV. 423, 423–24, 486–87 (2007) (suggesting, based on a tentative analysis, that pro se defendants achieve results at least as strong as appointed counsel).

to the performance of defense attorneys elsewhere in the state. The defense attorneys in District 13 together produced an average Value Added score of 2.06 in 32 cases initially charged as taking indecent liberties with a child. On the other hand, they generated an average score of -1.73 in 38 cases arraigned as a robbery with a dangerous weapon. An attorney presented with this report would want to inquire more closely as to why the results achieved in District 13 are so much stronger (relative to other defense attorneys in the state) for indecent liberties than for robbery. Perhaps the answer lies in local prosecutorial policies or the practices of local judges. It is also possible that defense attorneys in District 13 are failing to pursue a tactic in their robbery cases that works well elsewhere in the state. Whatever the explanations, the discrepancy deserves some attention among defense attorneys. Those who look for patterns in local outcomes know what questions to ask.

C. Customizing Reports

The design choices we made in our demonstration model of a rating system are not set in stone. Each aspect of the rating—the use of sentencing movement as the sole measure of attorney performance, the selection of a starting point sentence, and the choice of the relevant comparison pool of defense attorneys—could play out differently to produce a system quite distinct from ours. The ability to change these parameters is a strength in a rating system. A manager could customize the system to address questions that are local concerns or priorities.

In particular, a manager might opt for a multi-factor measure of attorney success. There are many aspects of attorney performance that our initial rating effort misses by concentrating only on cases that result in some amount of incarceration. It is a genuine success for an attorney when the state dismisses charges against a defendant, or the jury acquits at trial, or the judge imposes probation or some other community sanction that includes no period of detention.

The ACIS system includes data on each of these case-level events, and one could construct a measure of attorney success that includes them. Other relevant measures of attorney success,

such as the reduction of “embedded consequences” like immigration removal or loss of public housing, might become a part of the rating if the defense counsel organization were to collect its own data to supplement court data.⁹² If the defender office surveys its clients about their level of satisfaction with their lawyers, then that survey result could become a component of the attorney’s rating score.⁹³

A customized measure along these lines could take the form of an index score, assigning various percentages of an overall score to different types of success, blended to suit the needs of the rater. For example, the organization might construct a rating score that is based 30% on incarceration sentences imposed, 20% for non-incarceration sentences, 10% for embedded consequences, 10% for dismissals, 10% for acquittals, and 20% for client satisfaction. Alternatively, the manager might track each of these ratings separately, without attempting to combine them into a single rating.

V. *The Environment for Success*

With case-level success ratings in hand, an evaluator of defense attorneys might turn naturally to questions about causation. What factors explain why some attorneys achieve stronger scores than others? Are there recurring case features that make it harder for any attorney to succeed? And are there recurring aspects of an attorney’s background, experience, or work environment that are associated more often with stronger

92. For example, in *Padilla v. Kentucky*, the risk of deportation of an alien was a factor used in evaluating the performance prong of the analysis for ineffective assistance of counsel. See *Padilla v. Kentucky*, 130 S. Ct. 1473, 1483 (2010) (holding that counsel engaged in deficient performance by failing to advise defendant that his plea of guilty made him subject to automatic deportation).

93. Cf. Stephanos Bibas, *Rewarding Prosecutors for Performance*, 6 OHIO ST. J. CRIM. L. 441, 445 (2009) (proposing the use of surveys as one component of prosecutor evaluation). Defense, like prosecution, is multi-faceted. Every client is different, and some clients do not care primarily about the outcome in criminal court; they only want their day in court. Cf. Tyler, *supra* note 66, at 115–24. A rating that combined outcomes and client satisfaction might be closer to ideal.

sentencing movement for the client? In this Part, we illustrate some ways that performance data might combine with biographical details about individual attorneys and data about the local courthouse environment to inform evaluators as they think about the drivers of attorney success.

A. Who Are These Guys?

In an effort to answer these causation questions, we built a database for a subset of the attorneys in the statewide ACIS system. Our sample consists of 215 North Carolina criminal defense attorneys. We chose the sample based on two criteria: geographic diversity and the number of cases the attorneys handled. We selected 14 of the 44 prosecutorial districts from around the state (oversampling the busiest districts in the system), and for each of these districts, we identified the defense attorneys with the most case dispositions. For most districts, this meant selecting between 15 and 25 attorneys, but in a few districts, we selected fewer than 10, based on the relatively low volume of case dispositions in those districts. In some districts, the state operates a public defender's office; in others, the court appoints criminal defense lawyers from private practice. In both types of districts, defendants retain their own attorneys in some cases. Table 2 describes the 14 prosecutorial districts.

Table 2: Sample Districts for Attorney Background Study

Prosecutorial District	Cities/Counties	Public Defender Office?	Number of Attorneys Sampled
4	Jacksonville (Onslow, Duplin, Sampson, and Jones Counties)	No	20
5	Wilmington (New Hanover and Pender Counties)	Yes	27
9	Granville, Vance, Franklin, and Warren Counties	No	20
9A	Caswell and Person Counties	No	2
10	Raleigh (Wake County)	Yes	15
12	Fayetteville (Cumberland County)	Yes	31
13	Bladen, Brunswick, and Columbus Counties	No	18
18	Greensboro (Guilford County)	Yes	8
19D	Southern Pines (Moore County)	No	7
21	Winston-Salem (Forsyth County)	Yes	13
25	Burke, Caldwell, and Catawba Counties	No	18
26	Charlotte (Mecklenburg County)	Yes	10
27A	Gastonia (Gaston County)	Yes	22
29B	Henderson, Polk, and Transylvania Counties	Yes	4
Totals		8 Yes/ 6 No	215

Because we wanted to learn if one or more attributes of the defense attorney's background or experience could explain performance,⁹⁴ we collected data from public sources on a number of characteristics. These included personal characteristics such as gender, race (Caucasian, Black, Latino, Asian, Native American, or Other), and political party affiliation (Democrat, Republican, Libertarian, or Unaffiliated).

Second, we collected attorney characteristics that could reflect the quality of the lawyer's education and professional experience. These included the number of case dispositions, the year the attorney was admitted to the bar, the law school the attorney attended, the attorney's type of practice (classified as

94. See Ryan D. King, Kecia R. Johnson & Kelly McGeever, *Demography of the Legal Profession and Racial Disparities in Sentencing*, 44 LAW & SOC'Y REV. 1, 1, 26 (2010) (hypothesizing that racial and ethnic disparities in sentencing are mitigated where the legal profession is more diverse).

public defender, solo practitioner, or law firm), the size of the law firm,⁹⁵ whether the attorney limited his or her practice to criminal defense, and whether the North Carolina State Bar had ever disciplined the attorney.⁹⁶

Third, we gathered data that could shed light on the quality of the attorney's connections to the local community, to local jurors and law enforcement officers, and to prosecutors. In other negotiation contexts, longstanding relationships among the negotiating parties can predict strong performances.⁹⁷ Under this heading, we collected data about the primary city where the attorney practices, the city where the attorney lives, and whether the law school the attorney attended was located outside North Carolina.

The attorneys in the sample were predominantly male (78%) and white (75%). There were 16 African American attorneys, 2 Asians, 1 Native American, and no Latino attorneys. Years in practice ranged from 1 to 41, with a mean of 14.7 years and a median of 12 years. Thirteen attorneys had 1 year or less in practice. The number of cases handled ranged from 7 to 299, with a mean of 48.4 and median of 33. The attorneys in our sample attended 48 different law schools, most of them in-state schools.⁹⁸

95. While we believe the distinction between solo practitioners and law firm members is important, it is also true that the law firms where these attorneys worked were almost always quite small—usually less than five, and often just two or three. Fifty-two of the law firms consisted of five or fewer attorneys. In only six instances the law firm in question consisted of more than five attorneys; in only two instances the law firm in question consisted of more than ten attorneys.

96. For most of these characteristics, our data is close to complete; for other variables, such as the attorney's age, hometown, and prior service as an assistant district attorney, our research did not uncover enough data to make the variable useful in our analysis. Most of our information has been collected electronically, from a combination of public records such as voter registration lists and the list of licensed attorneys maintained by the North Carolina State Bar, as well as websites maintained by the attorneys themselves.

97. See Thomas B. Metzloff, Ralph A. Peeples & Catherine T. Harris, *Empirical Perspectives on Mediation and Malpractice*, 60 LAW & CONTEMP. PROBS. 107, 108 (1997) (discussing the notion that mediation is usually most appropriate when the parties involved have an interest in maintaining a long-term relationship).

98. The most common law schools attended were Campbell (54), UNC (31), and North Carolina Central (28), all of them in-state law schools. Most of the

There were 48 attorneys practicing as public defenders, 91 attorneys practicing privately but solo, and 63 attorneys practicing in a private law firm.⁹⁹ From information provided by the North Carolina State Board of Elections, we determined that 101 attorneys were registered as Democrats, 48 as Republicans, 35 as unaffiliated, and 1 as Libertarian. From information provided by the North Carolina State Bar's website, a surprisingly high number of the attorneys (12%) had been disciplined in some way, whether in the form of admonition, censure, reprimand, or suspension of license.¹⁰⁰

B. The Raw Materials and the Attorney Tools

Which of these biographical and system characteristics might affect the performance of defense attorneys? We begin by looking at the personal background of the attorneys. Unsurprisingly, there were no meaningful differences in the performance scores among attorneys based on gender or race.¹⁰¹ More interesting, the average score appeared at first glance to vary depending on the political affiliation of the defense attorney: Democrats produced an average score of -.13, Republicans a score of -.05, and unaffiliated attorneys registered an average of -.24, while attorneys with an unknown political affiliation produced positive scores.¹⁰² This apparent difference disappeared, however, when we used regression analysis to examine the relative importance of these personal background factors taken as a whole.¹⁰³

attorneys attended an in-state law school (135/191, 71%).

99. To determine type of practice, we sometimes had to rely on 2012 information. We were unable to determine type of practice for 11 attorneys.

100. We found disciplinary action for 25 of 208 attorneys.

101. Our ANOVA analyses for differences in Attorney Index Scores based on attorney gender, race, and age all produced insignificant results. For age, $p = .25$; for gender, $p = .56$; for race, $p = .99$.

102. Our ANOVA analysis for party affiliation showed a significant difference, with $p = 0.01$.

103. This ordinary least squares regression, which we called Model 1, took Attorney Index Score as the dependent variable. Independent variables were age, gender, and dummy variables for race and party affiliation. Model 1 was not significant, with $p = .366$ and $R\text{-squared} = .06$. This model uses attorney-level data; for a discussion of the use of case-level data, see *infra* note 117 and

Next, we explored factors relating to the “raw materials” available to the attorney in a given case: these included the average severity of the offense in the attorney’s portfolio of cases and the average level of the prior record among the defendants in the portfolio. The prosecutors and defense attorneys may treat serious crimes and extensive prior records systematically differently from lesser cases, which would affect how much movement in sentences one might expect from a defense attorney in that setting.

The raw materials variables also include the prosecutorial district where the attorney normally worked. The district attorney in one location might pursue distinctive policies, such as limits on the plea offers that assistant district attorneys are allowed to make for particular crimes. Similarly, the judges in one district might follow practices that limit the outcomes that are realistically available for a defense attorney to pursue, while judges in another district might allow the defense to negotiate a wider range of outcomes.

Each of these raw materials variables, standing alone, appears to have a significant effect on the attorney performance score.¹⁰⁴ When we add these variables to the regression model (controlling for the effects of any personal background variables), three of them make a significant difference on the performance score.¹⁰⁵ These include the offender’s prior record,¹⁰⁶ the

accompanying text.

104. For the ANOVA analyses of the differences in attorney performance scores among different offense levels, different offender prior record categories, and different prosecutorial districts, $p = .000$ in all three cases.

105. Model 2 retains all the variables from Model 1, and adds new independent variables: average offense class, average prior record category, and dummy variables for each of the prosecutorial districts, using District 5 as the reference category. The model is significant ($p = .000$) with an R-square statistic of .61 and an adjusted R-square of .55.

106. The offense level does not correlate with movement in the performance score at a significant level. Although offense seriousness intuitively would affect attorney performance, the score in our rating system is itself designed to account for the offense as charged. On the other hand, the score itself may not incorporate the full impact of the offender’s prior record. Even though the value added reflects a starting point that accounts for the offender’s prior record, our comparison to other attorneys considers all cases in the state with the same charged offense, regardless of prior record level.

attorney's age, and the presence of the attorney in the most distinctive prosecutorial district.¹⁰⁷

Finally, we turn to those features of the attorney's professional background and practice environment that might affect performance outcomes. One might consider these the "tools" that the attorney brings to bear on the raw materials that present themselves.¹⁰⁸

One of those tools involves the type of office where the attorney works: public defender office, solo private practice, or private law firm.¹⁰⁹ At first glance, there seem to be small differences among these practice settings. The average index score for public defenders was -.128; for attorneys in law firms, -.066; and for solo practitioners, -.004.¹¹⁰ The interaction of this feature with other aspects of the raw materials and the other attorney tools deserves further attention.

Another variable that initially attracted our attention is years of experience. Intuitively, it is difficult to know whether experience would move the performance score up or down. On the one hand, attorneys with more experience might learn from past efforts and become more effective over time for their clients.¹¹¹ On

107. The attorney's age correlates with years of experience, so we remove the age variable in Model 3. Prosecutorial District 21, which had a significant effect on attorney performance, claims the highest trial rate in the state, the highest use of the habitual felon law, and other distinctive features.

108. We recognize that some of the variables in Model 1, which we characterized as "personal background," could just as easily be treated as "tools" for the attorney to use for improved performance. One can easily imagine, for instance, that Hispanic attorneys (more likely to be bilingual) could be more effective with some clients; similarly, the attorney's gender might become a useful "tool" for certain types of offenses or offenders. Taken in the aggregate, any such advantages seem to balance each other out across all offense and offender types in our data.

109. Because we primarily analyze data at the attorney level for this study, we do not include a variable for "retained" versus "appointed" counsel. Attorneys in private practice in North Carolina typically represent clients of both types, and we make no effort here to categorize a private attorney as principally retained or appointed.

110. The median scores were -.029 for public defenders, .0086 for solo practitioners, and .029 for attorneys in private firms.

111. Some quantitative analysis finds such an effect in some locations. See Abrams & Yoon, *supra* note 74, at 1150 (finding that experienced attorneys achieve substantially more favorable outcomes for their clients than less

the other hand, attorneys might devote less energy to their cases after many years on the job and may invest less time in learning about any changes in criminal law and practice—particularly for complex changes in the law, such as the arrival of sentencing guidelines. Similarly, more experienced attorneys may over time come to accept local norms about acceptable outcomes and stop pressing so hard against that accepted courthouse culture.¹¹² This phenomenon might be called the “Sam Rayburn effect,” recalling the former Speaker of the House of Representatives’ advice: “If you want to get along, go along.”

The data from North Carolina offers tentative support for the latter unhappy story about experience. The average case performance score remains positive for the first 5 years in practice; the average case performance for almost every experience level above that is negative.¹¹³ Figure 1 summarizes the trend.

experienced attorneys).

112. Structured sentencing in North Carolina took effect in 1994. Structured Sentencing Act, N.C. GEN. STAT. §§ 15A-1340.10-23 (1994).

113. In Figure 1, the average index score for the category 0–5 years is based on 2393 cases and 45 attorneys; for 6–10 years, 2041 cases and 45 attorneys; for 11–15 years, 2225 cases and 42 attorneys; for 16–25 years, 1991 cases and 40 attorneys; and for 26 or more years, 1,847 cases and 42 attorneys.

Figure 1: Average Case-Level Performance Scores
by Attorney Experience Level



This performance difference based on experience of the attorney might reflect differences in the raw materials or practice setting of attorneys as they spend more years in practice. Again, the effect of this variable, after controlling for other influences, deserves close attention.

Finally, we were surprised to note that the attorney's criminal caseload (as measured by the number of sentenced cases that the attorney shepherded through the system) and an attorney's past discipline by the state bar for ethics violations did not correlate with the attorneys' average performance scores.¹¹⁴

We added these attorney tools variables to the attorney background and raw materials variables to learn more about the relative importance of these factors. Table 3 summarizes the results of this regression analysis.

114. The ANOVA for attorney caseload was not significant, with $p = .727$. It was also not significant for attorney discipline, with $p = .673$.

Table 3: Factors that Correlate with Attorney Performance Scores

Independent Variables	Standardized Coefficients	t	P
Personal Background			
Male	.019	.34	.74
Caucasian	.003	.05	.96
Unaffiliated w/Political Party	-.002	-.04	.97
Raw Materials			
Offense Class	.084	1.53	.13
Prior Record	.210	3.30	.00
Prosecutor District 4	-.023	-.30	.76
District 9	.085	1.22	.23
District 9A	.002	.03	.97
District 10	-.022	-.34	.74
District 12	.034	.44	.66
District 13	.022	.35	.73
District 18	.023	.35	.72
District 19D	-.095	-1.66	.10
District 21	-.655	-9.38	.00
District 25	-.035	-.50	.62
District 26	-.084	-1.51	.13
District 27A	.107	1.53	.13
District 29B	-.070	-1.29	.20
Attorney Tools			
Caseload	.016	.22	.83
Years in Practice	-.141	-2.56	.01
In-state Law School	-.016	-.29	.78
Bar Discipline	-.031	-.57	.57
Public Defender Office	-.092	-1.16	.25
Solo Practice	.080	1.26	.21
Constant		-2.73	.01
R-squared = .66	Adj. R-sqd = .60		
Model significance = .000			

This analysis is interesting both for what seems to matter and for what seems not to matter. The attorney's caseload, a history of bar discipline, attendance at an in-state law school, and a practice based in a public defender's office or a solo practice were all statistically insignificant.¹¹⁵ On the other hand, years of

115. Variations on Model 3 produced similar results. The results were consistent with those we report here when we included all dummy variables for race (with African-American attorneys as the reference category) and for political affiliation (treating Democrat as the reference category). When we removed all Prosecutor Districts, the defense attorney's base in a public

experience remained significant after controlling for other variables. Even though more experienced attorneys tend to handle cases with more serious charges and more extensive prior criminal history records, they still produce weaker results after controlling for those variables.

Among the raw materials variables, the offender's prior criminal record remained significant in this model, suggesting that attorneys who handle repeat felons tend to produce better results, all other things being equal. Perhaps these cases receive the full attention and efforts of the defense attorney because so much incarceration time is at stake; or maybe the prosecutors and high-volume defense attorneys in the state have reached a consensus that the sentencing grid assigns too much weight to prior record.

Finally, two of the prosecutorial districts produced distinctive results. Defense attorneys in Districts 21 and 19D tended to achieve less sentencing movement than attorneys in the comparison district.¹¹⁶ Other prosecutorial districts came close to significance in this model.¹¹⁷

The impact of these two prosecutorial districts flags an important limitation of this demonstration project. Our efforts to explain sentences (primarily negotiated outcomes) focus on a few features of the offense and the offender, along with a larger array of factors relating to the defense attorney's background, abilities, and resources. We have little information, however, on the professional background and resources of the other negotiating partner: the prosecutor. A defender's office that maintained

defender's office became significant ($p = .03$), but we believe this reflects that fact that public defender offices are located in only 8 of the 16 districts, and face distinctive prosecutor policies.

116. We treated Prosecutorial District 5 as the reference category for all of the district dummy variables.

117. We obtained similar but not identical results with a regression based on case-level data rather than attorney-level data. Because the case-level model uses an enormous number of records ($n = 10,645$), it produced significant results for the model as a whole, but an extremely low R-squared statistic (.03). In the case-level model, offense level proved to be a significant independent variable—it was reasonably close to significant in the attorney-level Model 3. The presence of the case in Prosecutor District 21 also showed up as a significant variable in this model.

biographical data about individual prosecutors in the district and accounted for structural features of the local prosecutor's office might be able to explain more of the performance of defense attorneys in single cases or across their entire portfolio of cases. An understanding of defense attorney success requires a detailed knowledge of prosecutors.

VI. The Uses and Users of Performance Scores

Based on this exercise with felony case data from the North Carolina courts, we conclude that it is possible to build a quantitative rating system for attorneys based on outcomes achieved for clients. These performance scores could become one component of an attorney evaluation. We now consider several responsible—and irresponsible—uses of performance scores, both by managers within defender organizations and by outsiders to those organizations.

A. Insider Uses

A wise manager of a defense attorney organization could use performance scores, together with other evaluation techniques, to improve the quality of services for clients and to get the most out of a limited budget. At the same time, a word of caution is in order. Quantitative measures of performance do tell us something, but it is important not to become mesmerized by the numbers. In the educational context, the evaluation of teachers based simply on annual movements in student test scores would be misleading. A school principal must combine test score information with personal observation of teachers in the classroom, review of lesson plans, and regular interaction with students and parents to get a realistic evaluation of the teacher.¹¹⁸ Relying too heavily on student test scores would be a destructive short cut.¹¹⁹

118. See Zorn, *supra* note 47 (arguing that a more comprehensive approach is needed to evaluate teachers rather than simply considering test scores).

119. See *id.* (noting that the National Academy of Sciences has twice spoken

The same is true for defense counsel and sentencing outcomes. The scores reveal something important about an attorney's effectiveness, but they cannot stand alone. Given the limited information that courts and defender organizations collect about any given case, a performance score based on an appropriately large number of cases tells the reader something about the attorney's entire portfolio, but not about the quality of the lawyer's work in any single case (or even in a small subset of cases). A manager of defense attorneys, like a school principal, must combine performance scores with personal observation, in-depth review of files in selected areas, and interaction with clients, judges, and prosecutors to get the full picture of an attorney's work.

Working within these limits, what personnel and operational decisions might benefit from the use of performance scores as one input? As we mentioned earlier, a manager could rely on regular reports of performance scores to identify attorneys who need closer monitoring. If that closer look seems to confirm the attorney's relatively weak performance score, some targeted training may be the next step. Weak performance scores might also point to a particular category of offense or offender that creates extra trouble for the attorney. The scores, combined with closer monitoring, could tell the manager the precise type of training that might be necessary. Again, if closer observation confirms the initial clues from the low performance scores, this insight might lead a manager to assign an attorney cases that play to his or her strengths.

Looking beyond the scores assigned to particular attorneys, a supervisor making case assignment decisions might draw insight from statistical analyses of raw materials and attorney tools. If studies of local conditions show that certain types of defense attorneys tend to get better results when they negotiate against certain types of prosecutors (say, younger defenders with local ties working across the table from younger prosecutors with local ties), then the supervisor making case assignments might try to benefit from this match-up. Much as a baseball manager knows

out about the dangers of basing decisions on erratic and inherently unreliable test scores filtered through imperfect and abstruse formulae).

which hitters on the bench have the best history against side-armed, left-handed relief pitchers, a supervisor in a defender organization should have more than intuition or anecdotal memory to inform her case assignment choices.

We also indicated in Part IV that a rating system could produce reports that focus on the combined outcomes that all attorneys in the office achieve, when compared to other attorneys in the state, for designated offense types. Where the office consistently performs above average, a manager might inquire more closely to learn more about what works well. For crimes where the entire office lags behind other defense attorneys in the state, the report could lead to closer inspection and reflection in the office. Perhaps the best response involves a shift of office resources from lower priority areas; perhaps more intensive training is the answer.

Performance ratings based on statewide data could also identify other regions where the defense bar obtains better results for a particular crime. A manager might explore the possible reasons for success in those other districts and then aim to replicate that success at home. If the defense success turns on conditions external to the defender organization itself (for instance, judicial practices might be the cause), the leadership of the defender organization could start lobbying other institutional actors for changes in their local practices.

Performance scores might also affect the hiring and compensation policies of a defender office. For instance, if an analysis of the attorney scores shows that attorneys with something between five and ten years of experience get the best results, all other things being equal, then the hiring strategy for the office might prioritize attorneys in that zone. If attorneys with more than twenty years of experience tend to produce weaker results, then the salary structure in the office might put less of a premium on experience.

B. Outsider Uses

While we see a wide range of possible uses for performance scores among managers and others working inside defender organizations, we close with a more skeptical take on the use of

these scores by other constituencies. The possibility of misinterpretation and misuse of the ratings by judges or potential clients is intolerably high. Those actors cannot follow up an initial numerical warning system with a more thorough individual evaluation.

Consider first a judge who might use performance scores. Suppose that the question of attorney competence arises during the adjudication of an ineffective assistance of counsel claim. One could imagine that an attorney's low scores in a performance rating system might qualify as relevant evidence for the petitioner; alternatively, the respondent might want to introduce evidence of the attorney's high scores to show that the attorney was generally effective.

Such judicial uses of a performance score would be misguided. The score tells us something about the results an attorney achieved in the past, compared to the results that other attorneys achieve in cases with at least some common characteristics. A score based on more prior cases increases our confidence that it was based on a normal distribution of cases, and that the unseen features of the attorney's cases balance each other out, for positive and for negative. In short, the score sheds light on the attorney's performance in an entire portfolio of cases. For many of the same reasons that rules of evidence severely restrict the use of prior similar acts to prove that a person committed a particular action, judges should resist the use of an attorney's portfolio performance score to prove that the attorney performed especially poorly (or well) in a particular case.

Similarly, one can imagine that a criminal defendant would want to consult attorney performance scores before hiring a lawyer. The score, however, would mislead potential clients because the assumption of random case distribution might not hold true for a given attorney score. The manager of a defender organization can account for this possibility by supplementing the performance score with other forms of monitoring, including personal observation, consultation with other observers of the attorney, and closer inspection of case documents. Potential clients do not have access to these alternative measures. For many of the same reasons that publication of teacher ratings based exclusively on student test performance tends to do more harm than good, we believe that publication of attorney

performance scores would not serve potential clients well. Although prospective clients have the most at stake in evaluating defense attorneys, they are also most likely to misunderstand ratings.

VII. Conclusion

The results-based measure of attorney effectiveness that we have constructed as a thought experiment for this Article could become something more than an academic exercise. It meets a widespread need in the field. When combined with other forms of observation and assessment, performance scores could enable defender organizations to improve their hiring, case assignment, compensation structures, and training to deliver better results for criminal defendants. Attorney ratings could become a common managerial tool among public defender organizations, or even in private law firms that specialize in criminal defense.

The ratings system we built here uses data currently available in many state court systems; it does not require an expensive layer of personnel to track the essential information about each case. The concept is flexible enough to adapt to local conditions, including the data available in the jurisdiction, and to place value on the aspects of attorney performance that the organization's leaders want to measure. While an office might build a better rating system by tracking more detailed case information, managers could gain some insight from a bare-bones system that tracks the amount of sentence movement in each case.

In the end, the barriers to the use of attorney ratings are cultural rather than technical or budgetary. While there are some defender organizations that use quantitative measures to track the activities of their attorneys collectively,¹²⁰ these managerial

120. See Interview with John Gross, Indigent Defense Counsel, Nat'l Ass'n of Criminal Defense Lawyers, at the Washington & Lee University School of Law Symposium: *Gideon* At 50: Reassessing the Right to Counsel (Nov. 9, 2012) (stating that defender organizations have begun to use performance evaluations—collective and individual—with a quantitative aspect in Virginia, Massachusetts, Texas, and San Mateo County, California).

habits cut against the grain. Defense organizations have a reputation for attracting people who are skeptical of hierarchy. Defender office leaders may very well reject a rating system that takes current prosecutor charging as a legitimate starting point for measuring attorney performance.

The legacy of *Gideon*—a world of thin resources to provide defense counsel across a wide swath of criminal defendants—might overpower these cultural barriers. With so many client needs and so few lawyers to meet them, managers could see immediate benefits from using performance measures to get more value out of limited budgets. Bureaucratic rationality may prove hard to resist, even in this corner of the world.