Safeguard or Barrier: An Empirical Examination of Bar Exam Cut Scores

Michael B. Frisby,* Sam C. Erman,** Victor D. Quintanilla***

Author Note

*Michael B. Frisby is a doctoral candidate in the Educational Foundations and Policy program at the University of Michigan School of Education. https://orcid.org/0000-0003-1475-0470. The authors gratefully acknowledge funding from AccessLex and contributions from Christina Chong-Nakatsuchi, Lindsey Sporrer, Jessica Stephan, Jacob Wasser, Mitch Winick, USC's law library, funding from AccessLex, and the Mindsets in Legal Education research team. Correspondence concerning this article should be addressed to Michael B. Frisby (mbfrisby@umich.edu) and Sam C. Erman (serman@law.usc.edu).

**Sam C. Erman, is a professor of law at the University of Southern California.

***Victor D. Quintanilla University Bicentennial Professor Professor of Law, Van Nolan Faculty Fellow Maurer School of Law, Co-Director Center for Law, Society & Culture, Affiliated Professor Department of Psychological and Brain Sciences, at Indiana University.

1

Abstract

In 2019 more than forty percent of aspiring law school graduates failed the bar exam. Nearly thirty thousand test-takers otherwise qualified to practice law were, given the score threshold required to pass the exam (the "cut score"), lost to the profession. Had the cut score been lower, many would now be lawyers. This exclusion disproportionately affects members of underrepresented and disadvantaged groups who stand to benefit most from entry into the legal profession. A common defense for retaining or raising cut scores is that doing so prevents lawyer malfeasance. But the bar exam is not designed for these purposes. This paper enters this scholarly and regulatory conversation by testing whether states' bar exam scores predict lawyer misconduct. If they do not, it would remove one argument against lowering bar exam cut scores to promote diversity and growth of the legal profession. Using data comprising states' bar exam cut scores and disciplinary records from the American Bar Association between 2013 and 2018, we employ statistical modeling to evaluate the relationship between cut scores and attorney discipline. We find no evidence that higher bar exam cut scores produce fewer complaints, charges, or disciplinary actions.

Keywords: discipline, malfeasance, bar exam, diversity

Introduction

In 2019, nearly seventy thousand people took the bar exam.[1] More than forty percent failed.[2] Given the existing scores required to pass those exams (the "cut score"), nearly thirty thousand test-takers otherwise qualified to practice law were lost to the profession.[3] Had the cut score been lower, many would now be lawyers.[4] So it goes every year, with staggering costs.[5] Legal educators devote substantial resources to teaching tens of thousands of people legal skills that never get put to use. A national crisis in access to justice grows more entrenched. Applicants invest three years and countless thousands of dollars of legal education, then find the way on the path they had charted to upward mobility and a professional career barred. The exclusion disproportionately affects the members of underrepresented and disadvantaged groups who stand to benefit most from entry. Concurrently, the profession's dire need to diversify goes unaddressed, perpetuating the lack of representation and inclusion for broad swathes of the public.

The reasons that the legal profession has chosen not to lower bar exam cut scores do it little credit. Legal regulators typically defend cut scores as measures of minimum competence, disparate racial impacts notwithstanding. But the bar exam has never been job validated, and fails to meet the substantive antidiscrimination standards imposed on most employment tests. This anomaly leads some critics to suggest that racism and anti-competition are the true drivers

---

[1] *2019 Statistics Snapshot*, The Bar Examiner (2020), https://thebarexaminer.org/2019-statistics/2019-statistics-snapshot/.

[2] *Id.*

[3] *Id*.

[4] Mitchel L. Winick et al., *Examining the California Cut Score: An Empirical Analysis of Minimum Competency, Public Protection, Disparate Impact, and National Standards*, AccessLex Institute Research Paper, Oct. 15, 2020, *available at* https://arc.accesslex.org/grantee/56/.

[5] *See, e.g.*, *2018 Statistics Snapshot*, The Bar Examiner (2020), https://thebarexaminer.org/statistics/2018-statistics/2018-statistics-snapshot/.

of heightened cut scores. More is expected from the profession entrusted with the rules for reducing discrimination, promoting equity, and ensuring fairness. A common defense for retaining or raising cut scores is that doing so prevents lawyer malfeasance. But the bar exam is not designed to weed out unethical people. Even if it accidentally predicted discipline, it could be inappropriate to use it for that purpose. And either way, use of the exam distracts attention from more effective, less discriminatory approaches, such as behavioral systems and regulations for practicing lawyers.

This paper enters this scholarly and regulatory conversation by testing whether lawyers' bar exam scores predict misconduct. If they do not, it would remove one argument against lowering bar exam cut scores to promote diversity and access to the legal profession. Importantly, the paper's aim is not to identify the best way to prevent lawyer misconduct; many better alternatives exist. It is instead a paper about bar exams, lawyer discipline, and the fundamental flaws of a particular strategy that limits diversity.

Reasonable Skepticism of Heightened Cut Scores

A robust scholarship justifies skepticism that heightened cut scores produce less dangerous attorneys. Bar exam advocates have long flown the banner of public protection in support of an instrument that excludes underrepresented populations. Yet, cut scores vary among jurisdictions and across time with no apparent empirical justification. The bar exam does not even purport to measure the traits and behaviors that most tend to result in findings of lawyer malfeasance. Prior empirical attempts to find relationships between bar exam performance and subsequent discipline have fared little better, given daunting methodological challenges.

The Exclusionary Backdrop of the Bar Exam

4

Whether measured by word or deed, exclusion was long the animating principle of the bar exam. At the inception of the modern, highly regulated system of entry into the legal profession, its architects cast racial exclusion *as* public protection. Consider U.S. Senator and American Bar Association President Elihu Root, who in 1916 confronted a bar with few Black, brown, and female members by inveighing[6] against the dangers of New York's ethnic-European bar:

> Fifty percent of the lawyers of this city are either foreign born or of foreign parents. And the great mass of them have in their blood . . . the traditions of the countries from which they came. . . . [T]his great mass . . . will change us unless we change them.[7]

Root led the ABA's efforts to erect the modern system of legal education and licensure that culminates with the bar exam.[8]

Skip forward half a century, and the bar exam still operated as an engine of exclusion. For technical reasons, the federal courts decided that Title VII's ban on racially discriminatory employment tests did not apply to bar exams.[9] But after every single one of forty Black applicants failed the Georgia bar exam in 1972, the United States Court of Appeals for the Fifth Circuit rejected the contention that exam passage established a "'minimal competence required to practice law.'"[10] Had Title VII applied, the court reasoned, the adverse racial impact and lack of professional validation study "would inexorably compel the conclusion that the examination" was illegal.[11]

---

[6] Elihu Root, Addresses on Government and Citizenship, 511-512, 519 (1916). Alba M. Edwards, Sixteenth Census of the United States: 1940 Population Comparative Occupation Statistics for the United States, 1870 to 1940, at 135, 165, 172 (1943).

[7] Root, *supra* note 6 at 479.

[8] *See, e.g.*, James P. White, *Legal Education in the Era of Change: Law School Autonomy*, 1987 Duke L.J. 292, 294-295.

[9] *See, e.g.*, *Griggs v. Duke Power Co.*, 401 U.S. 424 (1971); *Tyler v. Vickery*, 517 F.2d 1089, 1096 (1975).

[10] *Tyler*, 517 F.2d, at 1102, 1092 (apparently quoting the bar examiners).

[11] *Tyler*, 517 F.2d, at 1096.

Half a century later, exclusion and lack of access remain the norm. Less than 15% of today's U.S. lawyers are people of color, and low-income Americans (a disproportionately nonwhite population) receive adequate legal assistance for less than 15% of their civil legal problems.[12] The ABA, NCBE, and their state counterparts now have many programs to improve diversity and equality.[13] Yet, as we have shown elsewhere,[14] higher cut scores have substantial negative impacts on the diversity and ambit of a jurisdiction's legal profession.[15]

The Cost of Exclusion

This exclusion from practice of large numbers of law school graduates who are disproportionately people of color undermines fundamental commitments of the legal profession: justice, service, opportunity, public legitimacy, and fairness. Every otherwise qualified lawyer excluded from practice by a heightened cut score is one less attorney available to help close the

---

[12] American Bar Association, ABA National Lawyer Population Survey: 10-Year Trend in Lawyer Demographics: Year 2020 https://www.americanbar.org/content/dam/aba/administrative/market_research/national-lawyer-population-demographics-2010-2020.pdf (2020); Legal Services Corporation, *The Justice Gap: Measuring the Unmet Civil Legal Needs of Low-income Americans* 6 (2017), https://www.lsc.gov/sites/default/files/images/TheJusticeGap-FullReport.pdf.

[13] *See, e.g.*, *Diversity and Inclusion Center*, American Bar Association, https://www.americanbar.org/groups/diversity/ (Last accessed Oct. 24, 2020); *Resource Center for Access to Justice Initiatives,* American Bar Association, https://www.americanbar.org/groups/legal_aid_indigent_defense/resource_center_for_access_to_justice/ (last accessed Oct. 24, 2020); Victor D. Quintanilla et al., *Evaluating Productive Mindset Interventions that Promote Excellence on California's Bar Exam*, June 25, 2020, https://mindsetsinlegaleducation.com/wp-content/uploads/2020/10/MILE-ExecutiveSummary.pdf.

[14] *See also* Deborah J. Merritt et al. *Raising the Bar: A Social Science Critique of Recent Increases to Passing Scores on the Bar Exam*, 69 Univ. Cincinnati L. Rev. 929, 929-931 (2001). On perverse incentives of higher cut scores for legal education, *see id.*; Steven C. Bahls, *Standard Setting: The Impact of Higher Standards on the Quality of Legal Education*, 70 Bar Examiner 15 (2001).

[15] Bar abolitionists point to these racially disparate impacts as evidence that the profession should dispense with the bar exam altogether. *See, e.g.*, Edward F. Bell, *Do Bar Examinations Serve a Useful Purpose?* American Bar Association Journal, Dec. 1971, 1215. Our study is of primary relevance to a different question: if the bar exam is to be given, should the cut score be set high?

access-to-justice gap. The loss is multiplied given the disparate racial impact of higher cut scores. Attorneys of color are more likely than their white peers to enter work in government service, public service, or the public interest.[16] They typically also provide more services to clients of color, undertake more pro bono work, provide more mentoring to younger attorneys, and sit on more community organization boards.[17] Given pervasive racial inequities in U.S. life, the exclusion of aspiring attorneys of color from practice eliminates what would otherwise be an escalator to upward mobility and a professional career.[18]

The legitimacy of law as a central civic and governmental institution is at stake as well. As the Supreme Court explained in *Grutter v. Bollinger*, "it is necessary that the path to leadership be visibly open to talented and qualified individuals of every race and ethnicity" if the profession is to produce "leaders with legitimacy in the eyes of the citizenry."[19] Given that a key justification for heightened cut scores is the prevention of lawyer incompetence, additional norms of fairness would be violated were such scores neither designed to predict discipline nor shown to be predictive of it, as the literature suggests may be the case.

---

[16] Richard O. Lempert, *Michigan's Minority Graduates in Practice: The River Runs Through Law School*, 25 Law & Soc. Inquiry 395 (2000).
[17] Id.; Ronit Dinovitzer et al., *After the JD II: Second Results from a National Study of Legal Careers*, 72-73 (2009); Gita Z. Wilder, *Race and Ethnicity in the Legal Profession: Findings from the First Wave of the After the JD Study* 4-6, 15-16, 61-63 (2008); Michelle J. Anderson, *Legal Education Reform, Diversity, and Access to Justice*, 61 Rutgers L. Rev. 1011, 1015-19 (2009).
[18] On the increase in earnings by race associated with earning a J.D., see Frank McIntyre and Michael Simkovic, *Are Law Degrees as Valuable to Minorities*, 53 Intl. Rev. of L. and Econ. 23 (2018).
[19] 539 U.S. 306, 332 (2003); *see also* A. Leon Higginbotham, *The Case of the Missing Black Judges*, N.Y. Times, Jan. 29, 1992, A-21 ("Judicial homogeneity" makes "it difficult to have a court that . . . has the respect of most segments of the population," while "judicial pluralism breeds judicial legitimacy."); Brief of John Conyers et al as Amici Curae in Support of Respondents, Nos. 02-241, 02-516, *Grutter*, 539 U.S. 306 (Feb. 18, 2003), 5 (describing "core democratic values of full and fair political participation and responsive government for minority citizens").

The Case Against Cut Scores as a Measure of Subsequent Ethicality

Choices about cut scores display little rhyme or reason. As Gary Rosin observes, the choice of cut score "often has no empirical basis."[20] U.S. jurisdictions apply a wide range of cut scores and find them satisfactory. Law schools with similar scaled bar exam scores can have vastly different bar passage rates depending on the state they are in.[21] The phenomenon is particularly striking in California, where multiple non-ABA-accredited law schools have graduating classes with higher average scores on the Multistate Bar Exam than the average graduate of an ABA-accredited law school in the United States.[22] But because California has a heightened cut score and other states generally do not open their bar exams to applicants from schools only accredited by California, many of these above-average law school graduates cannot become lawyers.[23]

---

[20] Gary S. Rosin, *Unpacking the Bar: Of Cut Scores and Competence*, 32 J. Legal Profession 67, 72 (2008).

[21] *Id*. at 69.

[22] William Wesley Patton, *A Blueprint for a Fairer ABA Standard for Judging Law Graduates' Competence: How A Standard Based on Students' Scores in Relation to the National Mean MBE Score Properly Balances Consumer Safety with Increased Diversity in the Bar*, 24 Wash. & Lee J. Civ. Rts. & Soc. Just. 3 (2017).

[23] *Id.*; San Joaquin College of Law, Practicing Outside California, 2021, http://www.sjcl.edu/index.php/prospective-students/why-sjcl/practicing-outside-california. Many states permit graduates of California accredited law schools to sit for their bar exams if the applicant has already been admitted to practice in another jurisdiction. *Id*.

States regularly change their cut scores too, often for dubious reasons.[24] In the 1990s a third of states did so.[25] The vast majority moved cut scores upward, driving down bar passage rates even as applicants' quality and diversity rose.[26] Some saw anti-competitive practices at work; a second overlapping group perceived racism: "Why does a group of applicants that is one-fifth nonwhite have to show a higher level of competence than their mostly white predecessors displayed twenty years ago?"[27] Whatever the reasons, the results were stark: just as large numbers of talented nonwhite aspiring lawyers sought to enter the profession, legal regulators began erecting barriers that kept many out.

A further reason to be skeptical is that the bar exam is simply not designed to be an instrument that measures aspects of competence that would differentiate unethical attorneys from ethical ones or screen from practice those most likely to be disciplined. The problem is not that competence and ethics are unrelated; it is that the bar exam is a poor measure of either. Competence is generally understood to be the bundle of skills, attitudes, tendencies, abilities, pieces of knowledge, and the like that make for better or worse legal practice.[28] As a practical

---

[24] *See* Derek Muller, *A Few Longer Thoughts on the Four Debates about the Bar Exam*, Excess of Democracy, Aug. 22, 2017, https://excessofdemocracy.com/blog/2017/8/a-few-longer-thoughts-on-the-four-debates-about-the-bar-exam (describing how four states altered their cut scores in 2017 and how one state planned to raise it, then delayed doing so); *compare* Stephen Klein, *Setting Bar Exam Passing Scores and Standards*, The Bar Examiner, Nov. 2001, 12 (summarizing his methodology for helping states set cut scores) *with* Merritt et al., *supra* note 14 (critiquing of that method). Citing the lack of justification for cut scores that vary by jurisdiction, some have suggested that we move toward a uniform cut score. Joan Howarth, *The Case for a Uniform Cut Score*, 42 J. of Legal Prof. 69 (2017).

[25] Deborah Jones Merritt, *Raising the Bar: Limiting Entry to the Legal Profession*, 70 Bar Examiner 9, 11 n.4 (2001).

[26] *Id.*; Merritt et al., *supra* note 14, at 929-930, 937-939.

[27] Merritt, *supra* note 25, at 11; Benjamin Hoorn Barton, *Why Do We Regulate Lawyers: An Economic Analysis of the Justifications for Entry and Conduct Regulation*, 33 Ariz. St. L.J. 429, 433 (2001); *see also* Bahls, *supra* note 14, at 17 (proposing as the "most important test--whether practicing attorneys can pass the exam under actual exam conditions using a new cut score").

[28] Jeremy Cooper, *What Is Legal Competence?* 54 Modern L. Rev. 112 (1991); Maureen F. Fitzgerald, *Competence Revisited: Summary of Research on Lawyer Competence*, 13 J. Prof.

matter, competence can only be measured in broad strokes given the breadth and complexity of legal practice.[29] Were one to seek a comprehensive account, items could include: joy, grit, honesty, purpose, and professional pride, among others.[30] Just contemplating such an instrument illuminates the inadequacy of measuring competence writ large or or predicting ethics through a bar exam that primarily tests the memorization of subject matter knowledge, analysis, and recall under time-pressure.[31]

Legal Educ. 227 (1995); Deborah Jones Merritt & Logan Cornett, *Building a Better Bar* (Dec. 2020), https://iaals.du.edu/sites/default/files/documents/publications/building_a_better_bar.pdf.
[29] Cooper, *supra* note 28; Fitzgerald, *supra* note 28; Merritt & Cornett, *supra* note 28. Thus, a recent report by the Institute for the Advancement of the American Legal System that speaks just of the knowledge and skills necessary for the minimum professional competence necessary to begin practice lists the capacity for professional and ethical conduct; understanding of legal processes and sources and of threshold concepts in many subjects; identifying legal issues and clients' big picture concerns; researching and interpreting law; interacting with clients, colleagues, and others; managing and coping with workload; and self-directed learning. Merritt & Cornett, *supra* note 28; *see also* Cooper, *supra* note 28 (describing an earlier ALI-ABA report that reached similar conclusions). With such criteria in hand, it is possible to seek thresholds beyond which incompetence lies. This is the impulse that animates the bar exam's search for the minimum legal knowledge necessary for competent legal practice. *See, e.g.*, NCBE, *Bar Admissions During the COVID-19 Pandemic: Evaluating Options for the Class of 2020*, April 9, 2020, at 6. Other definitions display similar catholicism. *See, e.g.*, Fitzgerald, *supra* note 28 (reviewing post-1980 research on identifying competencies comprehensively and collecting pre-1980 studies).
[30] Roger C. Cranton, *Lawyer Competence and the Law Schools*, 4 Univ. Ark. Little Rock L.J. 1, 8-9 (1981). If competence were ever perfectly measured, capacity and performance would merge. Lawyers who achieved higher scores would be superior attorneys. Those judged incompetent would perform below minimum standards, Fitzgerald, *supra* note 28, at 248; Cooper, *supra* note 28, at 113-114, and thereby frequently violate ethical standards, given that the Rules of Professional Conduct demand minimum standards of performance, *see, e.g.,* ABA Model Rules of Professional Conduct §§ 1.1, 1.3. Such precise, predictive measurement is impossible, of course.
[31] On what the bar exam tests and on actual causes of lawyer discipline, *see* Judith Welch Wegner, *Contemplating Competence: Three Meditations*, 50 Valparaiso Univ. L. Rev. 675, 714 (2016); Susan R. Martyn, *Lawyer Competence and Lawyer Discipline: Beyond the Bar*, 69 Georgetown L.J. 705, 723 (1981); Norman Krivosha, *Lawyer Competence from Another Perspective*, 68 Am. Bar Assoc. J. 828, 828-829 (1982); Cranton, *supra* note 30, at 7-9; Raymond Marks & Darlene Cathcart, *Discipline Within the Le- gal Profession: Is It Self-Regulation?* 1974 U. Ill. L.F. 193, 212 (finding that lawyer neglect was the principal complaint filed by clients with an attorney grievance committee). *But see* Jeffrey S. Kinsler, *Is Bar Exam Failure a Harbinger of Professional Discipline?* 92 St. John's L. Rev. 883,884-891 (2017)

But even if the choice of a heightened cut score accidentally predicted discipline, it could

be inappropriate to choose one for that purpose. Leslie Levin and colleagues draw such a lesson

from character and fitness investigations, which collect predictors of subsequent discipline that

do not get used as reasons to deny permission to practice.[32] For instance, applicants report

whether they are men and whether they have previously defaulted on a student loan. Both groups

are more likely to be disciplined than women and those without such defaults.[33] Yet all are

permitted to practice law if otherwise qualified.[34] One reason is the relative lack of discipline

within the legal profession, which renders even these statistically significant predictors of

subsequent discipline trivial because they lack a meaningful effect size.[35] Another reason is the

general rule that bad acts are proper bases of punishment, but that bad propensities are not.[36]

These aspiring lawyers can claim that through hard work, money invested, expertise attained,

and forgone opportunities, they have earned the opportunity to practice.

The unfairness of heightened cut scores is exacerbated because they often represent

daunting and unequally distributed barriers to entry. Like other high-stakes tests such as the

SAT, GRE, and LSAT, the bar exam reproduces and compounds discrimination by

---

(hypothesizing that bar exam failure is a good measure of traits that it is not designed to test: lack of diligence and incompetent performance). Conversely, the bar exam omits measurement of many traits and behaviors that are central to professional success. *See* Marjorie M. Shultz and Sheldon Zedeck, *Predicting Lawyer Effectiveness: Broadening the Basis for Law School Admission Decisions*, 36 L. & Social Inq. 620, 625 (2011).

[32] LESLIE C. LEVIN ET AL., A STUDY OF THE RELATIONSHIP BETWEEN BAR ADMISSIONS DATA AND SUBSEQUENT LAWYER DISCIPLINE 1 (Mar.15, 2013), https://perma.cc/97RX-GYQK; Leslie C. Levin et al., *The Questionable Character of the Bar's Character and Fitness Inquiry*, 40 L. Soc. Inquiry 51 (2015).

[33] LEVIN ET AL., *supra* note 32; Deborah J. Merritt, *Bar Exam Scores and Lawyer Discipline*, Law School Cafe (blog). June 3, 2017. https://www.lawschoolcafe.org/2017/06/03/bar-exam-scores-and-lawyer-discipline/.

[34] LEVIN ET AL., *supra* note 32; Merritt, *supra* note 33.

[35] LEVIN ET AL., *supra* note 32.

[36] Merritt, *supra* note 33.

disproportionately excluding members of racial and ethnic minority groups.[37] High-stakes exams are typically offered in contexts that raise worries among test takers that underperformance would confirm negative stereotypes about the intellectual capacity of groups to which they belong. This "stereotype threat" tracks its animating stereotypes in disproportionately burdening test takers of color.[38] As we have shown in other work, legal education is also rife with high-stress, low-belonging, fixed-mindset contexts that particularly harm aspiring lawyers of color.[39] Add to that the enormous debt owed to U.S. communities of color who continue to be systematically denied access to education, well-paying jobs, sociopolitical participation, and other essential resources for success.[40] The result is a profession that misses an opportunity to promote diversity, equity, inclusion, and access to justice out of suspicion toward aspirants who have otherwise proven themselves by successfully graduating from a college or university,

---

[37] Kelly Ochs Rosinger et al., *The Role of Selective College Admissions Criteria in Interrupting or Reproducing Racial and Economic Inequities*, 92 J. Higher Educ. (2021); Casey W. Miller et al., *Typical Physics Ph.D. Admissions Criteria Limit Access to Underrepresented Groups But Fail To Predict Doctoral Completion*, 5 Science Advances 7550 (2019); Aaron N. Taylor, *The Marginalization of Black Aspiring Lawyers*, 13 FIU L. Rev. 489 (2019); Case Miller & Keivan Stassun, *A Test That Fails,* 510 Nature Careers 303 (2014)

[38] *See* Claude M. Steele & Joshua Aronson, *Stereotype Threat and the Intellectual Test Performance of African Americans*, 69 J. PERS. SOC. PSYCHOL. 797 (1995); Sian Beilock et al., *Stereotype Threat and Working Memory: Mechanisms, Alleviation, and Spillover*, 136 J. EXPERIMENTAL PSYCHOL. 256 (2007); Toni Schmader et al., *An Integrated Process Model of Stereotype Threat Effects on Performance*, 115 PSYCHOL. Rev. 336 (2008); Steven Spencer et al., *Stereotype Threat*, 67 ANN. REV. PSYCHOL. 415 (2016).

[39] Dorainne J. Green et al., *Group-Based Inequalities in Relationships in Law School Predict Disparities in Belonging, Satisfaction, and Achievement in Law School* (revise and resubmit received J. OF EDUCATIONAL PSYCHOLOGY); Victor D. Quintanilla & Sam Erman, *Mindsets in Legal Education*, Journal of Legal Education (forthcoming 2021). Studying for the bar exam involves many costs, including expensive bar preparation courses, large investments of time, and powerful psychological headwinds. All of these could be mitigated by lower cut scores.

[40] G. Ladson-Billings, *From the Achievement Gap to the Education Debt: Understanding Achievement in U.S. Schools*, 35 Educational Researcher 3 (2006).

matriculating into law school and to commencement, accomplishing these milestones and many more.

Empirical Challenges to Studying Lawyer Discipline

Prior empirical attempts to establish whether heightened cut scores reduce rates of ethical lapses among lawyers have faced four noteworthy challenges: data shortcomings, lag times, unobserved distribution, and proxies that may be confounding variables. The data problem arises because most scholars base their analyses on the disciplinary statistics that legal regulators maintain and that the ABA aggregates: complaints, charges, and discipline against attorneys.[41] There are compelling reasons to use this data. The information is accessible, can be compared across jurisdictions, and reflects the judgment of the profession as to what counts as objectionable.[42] Indeed, this study uses this data because there is, in fact, no better data for this investigation. Like most data, however, it presents challenges. Given consumers' lack of legal training,[43] their complaints may measure deficits in lawyers' social skills better than they do lawyers' performance at legal tasks.[44] Charging decisions that do not result in discipline are noisy signals. Even discipline decisions are heavily mediated by who complains, by the state's investigatory capacity, and by the types of failings subject to discipline.[45] The degree of

---

[41] Robert Anderson & Derek T. Muller, *The High Cost of Lowering the Bar*, 32 Georgetown J. Legal Ethics 307 2017; Kinsler, *supra* note 31; LEVIN ET AL., *supra* note 32; Levin et al., *supra* note 32.

[42] *See, e.g.*, ABA, Survey on Lawyer Discipline Systems 2018, https://www.americanbar.org/groups/professional_responsibility/resources/surveyonlawyerdisciplinesystems2014/.

[43] Cooper, *supra* note 28, at 116-117.

[44] A related literature in the medical context finds that doctor apologies reduce the risk of malpractice litigation. *See, e.g.*, B. Ho & E. Liu, *Does Sorry Work? The Impact of Apology Laws on Medical Malpractice*, 43 J. Risk & Uncertainty 141 (2011).

[45] Merritt, *supra* note 33; Charts VII, IX, in ABA, *supra* note 42; ABA, Jurisdictional Rules Comparison Charts, visited Feb. 11, 2021, https://www.americanbar.org/groups/professional_responsibility/policy/rule_charts/.

influence these issues have is unclear. Such problems are not unique to this study, and will persist in all future studies using this data so long as oversight improvements, funding for a robust investigative capacity, and a willingness to discipline the full range of lawyerly incompetence are not forthcoming. A profession committed to improving lawyer competence and deterring incompetence would not rely so heavily on the bar exam for that purpose.

Lag presents another problem because discipline of lawyers during their first ten years of practice is rare.[46] Indeed, rates of discipline accumulate so slowly such that by the *35th* year of practice only 5% of lawyers have faced some form of discipline; 95% of lawyers retain spotless disciplinary records.[47] Any systematic causal link between cut scores and discipline would require that bar exam underperformance prior to starting practice predicted misconduct into the twilight of a professional's career -- going off like an intergenerational time bomb after doing no measured damages for upwards of three decades.[48]

Jeffrey Kinsler sought to address the problems of lag between bar passage and discipline by focusing on the relationship that he found between prior bar exam failure and early-career malfeasance in Tennessee.[49] But his examination of discipline in 2005-2016 of lawyers who had passed the bar in 2005-2014 identified just *fourteen* attorneys who both previously failed the exam and subsequently received discipline--hardly a crisis demanding a response.[50] Tennessee's

---

[46] Anderson & Muller, *supra* note 41; Kinsler, *supra* note 31; Merritt, *supra* note 33.

[47] Anderson & Muller, *supra* note 41.

[48] *See* Merritt, *supra* note 33.

[49] Kinsler, *supra* note 31.

[50] Kinsler, *supra* note 31. Fourteen is not reported, but can be calculated from the following reported data: 7,256 lawyers passed the Tennessee bar exam during the relevant period; 69 of them were disciplined during the period, 281 of the lawyers who passed during the period failed the bar more than twice; 8.37% of those who passed, did so on the second attempt; 87.76% did so on the first attempt; among those who passed on the first attempt, the discipline rate was 0.864%. *Id.*, at 894, 897. The approach means that the reported discipline rates have more digits of accuracy than does the population of those disciplined.

cut score was generally well below the national median when this non-crisis occurred.[51] Because

such early-career discipline is quite unusual, it is also unclear what lessons can be drawn for the

larger problem of later-career disciplinary infractions.

A third problem is that intra-jurisdiction studies can only examine the discipline rates of

lawyers who pass the bar exam, which at best, renders speculative any conclusions about those

who would have passed the exam at a lower cut score.[52] Changing a cut score may also alter how

people study for the bar exam and thus how they perform on it.[53]

Finally, some studies use proxies for bar exam scores: prior bar exam failure or law

school rank.[54] But law school rank, rather than bar exam performance, is probably driving the

---

[51] Tennessee had a cut score of 1250 through 2010, and a cut score of 1350 beginning in 2011.
National Conference of Bar Examiners and American Bar Association Section of Legal
Education and Admissions to the Bar, Comprehensive Guide to Bar Admissions Requirements
[2000-2014] (2000-2014). The median cut score used by states during these years was 135,
except that before 2011 the median was sometimes a point or two lower. *Id.*

[52] *See* Derek Muller, *High-Level Implications: California Supreme Court Reduces Bar Exam Cut
Score from 144 to 139*, Excess of Democracy (blog), July 16, 2020,
https://excessofdemocracy.com/blog/2020/7/high-level-implications-california-supreme-court-
reduces-bar-exam-cut-score-from-144-to-139 (issuing a down-to-the-percentage-point prediction
as to the lifetime discipline rates of those attorneys who will be admitted to the California bar as
a result of the cut score being lowered from 1440 to 1390).

[53] Muller sidesteps such concerns with the predicated "all else being equal." Muller, *supra* note
52.

[54] Anderson & Muller, *supra* note 41 (using law school as a proxy for LSAT score as a proxy for
MBE score as a proxy for overall bar exam score); LEVIN ET AL., *supra* note 32, at 21, 28;
Kinsler, *supra* note 31; Merritt, *supra* note 33. Anderson & Muller, *supra* note 41, also identifies
a relationship between taking the bar in February and being subject to discipline subsequently.
This result, which matches what other researchers have found, *see* Merritt, *supra* note 33, is
largely derivative of the relationships between law school rank and subsequent discipline and
between having repeated the bar exam prior to passage and subsequent passage. Students at
lower-ranked law schools are more likely to engage in part-time study and thus more likely to
graduate and to take the bar exam off cycle. *See* ABA, 2009-2013 Total Part-Time JD
Enrollment by Gender and Ethnicity, visited Feb. 11, 2021,
https://www.americanbar.org/groups/legal_education/resources/statistics/statistics-archives/;
compare, e.g., State Bar of California, General Statistics Report July 2019 California Bar
Examination, Dec. 20, 2019, 1, http://www.calbar.ca.gov/Portals/0/documents/July2019-CBX-
Statistics.pdf? [hereinafter July 2019 Cal. Bar Stats], *with* State Bar of California, General
Statistics Report February 2020 California Bar Examination, June 26, 2020, 1,

result in both cases. Whereas the bar exam is not designed to measure the traits most associated with lawyer discipline, one's law school influences one's career trajectory in ways that shape one's susceptibility to subsequent discipline.[55] Unlike bar exam scores, one's law school appears prominently on one's resume and so provides a ready basis for discrimination. Unsurprisingly, graduates of higher-ranked law schools are more likely than graduates of lower-ranked ones to secure coveted big-firm jobs and slots in prosecutors' offices.[56] Graduates of lower-ranked schools gravitate to small and solo firms where over 90% of disciplinary sanctions are imposed.[57]

Lower rates of discipline of big firm lawyers and prosecutors may have more to do with the nature of their practice than with their competence upon entering into legal practice. Many law firms and prosecutor offices leverage and reinforce their elite status by ensuring that their attorneys' malfeasance does not result in complaints, charges, or discipline.[58] Law firms cultivate repeat-player clients who consequently have more opportunities to settle disputes

---

https://www.calbar.ca.gov/Portals/0/documents/FEB2020-CBX-Statistics.pdf [hereinafter Feb. 2020 Cal. Bar Stats]. Most people who fail the bar exam first take it in July, so repeaters are disproportionately present in February. *Compare, e.g.*, July 2019 Cal. Bar Stats, *supra*, *with* Feb. 2020 Cal. Bar Stats, *supra*.

[55] Levin et al., *supra* note 32; LEVIN ET AL., *supra* note 32.

[56] Merritt, *supra* note 33; LEVIN ET AL., *supra* note 32, at 29.

[57] Merritt, *supra* note 33; *see also* William Wesley Patton, *A Rebuttal to Kinsler's and to Anderson and Muller's Studies on the Purported Relationship Between Bar Passage Rates and Attorney Discipline*, 93 St. John's L. Rev. 43 (2019); LEVIN ET AL., *supra* note 32, at 29; Levin et al., *supra* note 32, at 56. Indeed, this relationship is part of a self-reinforcing cycle of disadvantage. Top-tier law schools primarily enroll advantaged students. Lower-tier law schools enroll many more members of disadvantaged groups. Then, the mostly advantaged students from top-tier law schools get routed into firm jobs and prosecutors' offices, where formal discipline rarely lies. Lawyers form disadvantaged groups tend to hail from lower-rank law schools, whom the profession routes into the small and solo firms upon whose attorneys the professional imposes nearly all its discipline. *See* Taylor, *supra* note 37.

[58] Patton, *supra* note 57.

bilaterally.[59] Larger firms also have the resources to create ethical infrastructure, which reduces neglect and associated complaints.[60] By contrast, though prosecutorial misconduct is rampant in some jurisdictions, friendly doctrines and power imbalances generally guarantee impunity.[61]

The nature of solo and small-firm practice also explains why such lawyers receive more complaints (and the charges and discipline that follow). Many lack adequate office support, which can lead to neglect of client matters and failures to return phone calls.[62] Such lawyers also occupy lower-status niches in the legal profession, making them subject to legal-regulators' bias.[63] Their practices involve more one-off, personal-plight representations with vulnerable and emotionally invested clients who have little recourse outside the disciplinary process.[64] Though lawyers at smaller practices are more likely to have cash flow problems and greater personal control over client funds, they do not disproportionately steal from client funds or engage in similarly serious misconduct.[65]

An irony lurks here. The broad policy question to be answered is whether the gains in the diversity, inclusion, representation, and capacity of the legal profession that would flow from lower cut scores will benefit the public overall. A common dissenting claim is that lowering cut scores could result in dramatically higher rates of discipline, thereby diminishing the quality of public protection. Studying just that question tends to focus public attention on speculative and, at most, modest harms of lowering cut scores rather than on its large, demonstrable benefits.

---

[59] Levin et al., *supra* note 32; David B. Wilkins, *Who Should Regulate Lawyers?* 105 Harv. L. Rev. 801 (1992).

[60] Levin et al., *supra* note 32, at 56; Patton, *supra* note 57.

[61] Patton, *supra* note 57.

[62] Patton, *supra* note 57; LEVIN ET AL., *supra* note 32, at 29, 37; Levin et al., *supra* note 32, 56.

[63] Levin et al., *supra* note 32, at 56; Merritt, *supra* note 33.

[64] Levin et al., *supra* note 32.

[65] Merritt, *supra* note 33; LEVIN ET AL., *supra* note 32, at 29, 37.

Such temporary blinders might be justified if they sharpened the part of the inquiry that was their focus. Instead, the effect has been to distort without clarifying.

Current Study

The current study seeks to evaluate the extent to which the cut score used as the passing score of the bar exam corresponds to improvement in a variety of public protection measures. We hypothesize that higher bar exam cut scores will be inert with respect to (1) decreasing the number of complaints filed against attorneys by the public, (2) decreasing the number of charges filed against attorneys, and (3) decreasing the number of disciplinary actions taken against attorneys. Said another way, we believe that there will be no evidence that suggests higher cut scores produces fewer complaints, charges, and disciplinary actions against attorneys. To evaluate these hypotheses we employ statistical modeling to the combined disciplinary records from the American Bar Association (ABA) and states' cut scores from 2013 to 2018.

<div align="center">

Methods

</div>

Sample

Disciplinary data, which consists of complaints brought by the public against attorneys, charges filed after probable cause, disciplinary actions taken against attorneys (henceforth collectively referred to as public protection data), as well as the number of active attorneys, are derived from the Survey on Lawyer Discipline Systems (SOLD), administered and maintained by the ABA's Center for Professional Responsibility. According to the ABA, SOLD data is intended to educate the public, the profession, the news media, courts and disciplinary agencies about sanctions imposed, caseload, budget, and staffing activities in each jurisdiction.

We collected the number of complaints filed, charges filed, disciplinary actions, and the number of active attorneys for each available state from 2013 until the most recent SOLD year of

<div align="center">

18

</div>

2018. These multi-jurisdictional records come from up to 48 U.S. jurisdictions, with the precise number of states reporting to the ABA varying by year. Additionally, several states were not accounted for in these ABA reports for particular years. In pursuit of exhaustive analyses, we gathered discipline data for these states from official reports on their respective state websites. These supplemental states included California 2013-2018, Massachusetts 2016-2017, Missouri 2013, Montana 2017, Nevada 2015-2016, New Hampshire 2013, Ohio 2014, and South Carolina 2013 & 2017. For future analyses, we refer to these as supplemental states.

All public protection measures were converted into counts per 1000 attorneys by multiplying the total incidences by 1,000 and dividing by the number of attorneys. For example, Alabama had 13,754 active attorneys in 2016 and 1,149 complaints received by a disciplinary agency. This is converted to (1,149 x 1,000)/13,754 = 83.5 complaints per 1000 attorneys. These modified variables were used as outcomes in statistical analyses.

Our cut score data consists of every state's minimum passing bar exam score dating from 2013-2018. Since 1994, the NCBE has released annual comprehensive guides to bar admission requirements, which include the minimum passing score for each state. We have compiled these reports to determine the minimum bar exam cut score required for licensure by each state between 2013 and 2018.

Measures

Complaints Brought by the Public Against Attorneys

We used the SOLD's records of complaints received each year by the state's disciplinary agency. Per the ABA, complaints include any information received by the disciplinary agency regarding lawyer conduct that requires a determination as to whether the disciplinary agency has jurisdiction over the lawyer or matter(s) complained of, or whether sufficient facts are alleged

19

that would, if true, constitute misconduct. Notably, if complaints were handled separately by a central intake or consumer assistance program, they were not counted in our measure. Complaint counts from central intake or consumer assistance programs were unavailable for most states, even those that reported the use of such programs. Additionally, the SOLD reports complaints pending from prior years, complaints summarily dismissed or screened out, complaints investigated, and complaints dismissed after investigation. Because these additional measures imply action taken on complaints, we felt they were unrelated to the posited relationship between cutscore and complaints *filed*, and thereby chose to discard this additional data.

Charges Filed After Probable Cause

Our use of charges corresponds to the SOLD's record of lawyers charged after probable cause determination. The ABA defines charges this way:

> After a determination has been made that there is probable cause to believe that misconduct occurred, any document, pleading or notice filed by the disciplinary agency or appropriate authority with the designated adjudicatory tribunal, wherein a lawyer is charged with specified acts of misconduct and violations of the rules of professional conduct and a disciplinary sanction is sought.

We collected the number of charges for each state per year. Notably, charges are not a prerequisite for discipline in all states, and thus some cases arise in which the number of disciplinary actions is greater than the number of charges.

Disciplinary Action

Disciplinary action consisted of two types: private and public. Private discipline includes action such as admonition, reprimand, or letter of warning/caution. Public discipline includes involuntary disbarment, disbarment on consent, suspension (excluding interim suspension), interim suspension (for risk of harm or criminal conviction), admonishment, reprimand, censure, probation, an order to pay restitution, and an order to pay costs. The SOLD provides the overall

number of public and private disciplinary actions taken for each state each year, in addition to the total number for each specific disciplinary measure.

Arizona, District of Columbia, Florida, Illinois, Kansas, Maryland, New Hampshire, New Jersey, Ohio, Oregon, and West Virginia do not have a public/private distinction. In these cases, data was only listed as public discipline (i.e., no private discipline was indicated). Additionally, SOLD records included notes from some states indicating, for example, that letters of caution or warning are not considered disciplinary actions. The full set of notes are relatively few and can be reviewed in the ABA's official SOLD report. We have accounted for these details in our series of analyses.

We harnessed the total number of public and private disciplinary actions taken against attorneys as our primary outcome measure. That is, we summed the total number of private and public disciplinary actions provided by the SOLD.[66] Like the other outcome measures, this outcome was converted to total disciplinary actions per 1000 attorneys. As a secondary set of analyses, we also looked at each specific disciplinary outcome (e.g., disbarment) with and without data that included states' notes as caveats.

Data analysis plan

Given the research questions' emphasis on generic relationships between bar exam cut scores and complaints, charges, and disciplinary actions taken against attorneys, we took a multimodal approach to data analysis that would allow us to uncover different ways in which these variables may relate to one another. Between 2013 and 2018, each state would provide annual counts for each public protection measure, totalling approximately 275 observations across this time window (the actual number will vary by outcome and analytic approach; e.g., see

---

[66] *See* columns 10 and 11 of the SOLD annual report.

Electronic copy available at: https://ssrn.com/abstract=3793272

Table 1). First, we used linear regression to explore the relationship between cut scores and public protection measures. Second, we aggregated the data by taking the mean for each state and reanalyzed the data using linear regression. This was done to remove state-related variation from the data and to only examine relationships between cut scores and disciplinary averages. Third, we applied multilevel modeling to control for within-state variation. Collectively, these methods offer a more comprehensive look at the relationships between cut score and attorney discipline than any single approach could offer, and therefore, allows for more robust statements to be made concerning their relations.[67] Each method is briefly introduced below.

Linear Regression (Disaggregated)

Linear regression is arguably the most well-known and employed method in statistical analysis. Ordinary least squares (OLS) regression is one approach to linear regression. We use OLS regression to model the relationship between the cut score required to pass the bar exam and the number of complaints, charges, and disciplinary actions taken in each state from 2013-2018. Given the data, OLS regression offers the best *linear* fit to the data. A linear fit is monotonic, thereby addressing the question of whether increases in the bar exam cut score correspond to increases (or decreases) in the number of complaints, charges, and/or disciplinary actions taken against attorneys. Linear regression also provides useful statistics such as effect

---

[67] Additional models were fit to this data, but are not reported here. Most notably, we fit polynomial (quadratic and cubic) models to the data, but chose to omit them from the report for three reasons. First, our primary aim was to evaluate the claim that higher cut scores increase public protection by decreasing the volume of complaints, charges, and discipline. Quadratic and cubic fits allow this relationship to wax and wane, thereby obscuring our ability to address this aim. Second, in most cases the quadratic and cubic fits showed little improvement to the explained variance, which was maximized at around 8%. Third, visually these fits appeared to overfit the data. Without a validation data set to confirm these fits, endorsement of overfitted models could suggest spurious conclusions. By focusing on the three models identified in this manuscript, we feel we have remained focused on the research question.

sizes and the proportion of variance explained (R^2), each of which offers insight into the strength of a given relationship.

Linear Regression (Aggregated)

In this series of models, we calculated each state's average public, private, and total public protection measures, and used this average as the outcome variable rather than the disaggregated data. Thus, there is only one outcome for each state, as opposed to one outcome for each year. Thereafter, the analytic approach was identical to the former, disaggregated data: OLS linear regression. This approach distills the within-state variation into a single estimate of each public protection measure for that state. By virtue of being the state *average*, it has some appeal of being closer to what we would expect from that state, and correspondingly serves as a better estimate of an average observation at a given cut score. However, despite its intuitive appeal, this approach may needlessly sacrifice statistical power to detect relationships between cut score and public protection. Multilevel modeling was applied as a follow up analysis.

Multilevel Modeling

Multilevel modeling (also called mixed-effects modeling) is a statistical method which allows for the control of nested data.[68] Nested data happens whenever there are multiple observations within a single unit of analysis, and multiple units exist within the data. Analyzing school data, for example, may include 20 students per classroom, and 50 classrooms. Mathematically, a baseline model with no predictors that controls for classroom effects would be written as

$$Y_{ij} = \gamma_{00} + u_{0j} + e_{ij},$$

---

[68] B.West, Linear Mixed Models (2015); J. Hox & J. Roberts eds., Handbook of Advanced Multilevel Analysis (2011).

Electronic copy available at: https://ssrn.com/abstract=3793272

where i represents the ith observation (the student), and j represents the jth cluster (the classroom). The statistic γ_{00} represents the overall mean of the dependent measure Y, while u_{0j} captures the classroom specific adjustment to γ_{00} (referred to as the cluster-level error term). The final term e_{ij} captures the deviation of observation i from its cluster mean (the deviation of a given student from her classroom average).

When predictors are added to the model, the expression is often generalized to the following by using matrix notation.

$$\mathbf{Y_j} = \mathbf{X_j\gamma} + \mathbf{Z_jU_j} + \mathbf{e_j},$$

where $\mathbf{Y_j}$ is now an $n_j \times 1$ response vector for the jth cluster, $\mathbf{X_j}$ is an $n \times p$ design matrix; $\mathbf{\gamma}$ is an unknown $p \times 1$ vector of fixed parameters to be estimated; $\mathbf{Z_j}$ is an $n_j \times k$ design matrix of random effects; $\mathbf{U_j}$ is a $k \times 1$ vector of unknown random effects to be estimated; and $\mathbf{e_j}$ is an $n_j \times 1$ vector of residuals. Newly incorporated predictors are included in the $\mathbf{X_j\gamma}$ piece of the expression, and new incorporated cluster specific adjustments such as u_{0j} are incorporated into the $\mathbf{Z_jU_j}$ part of the expression. The vector $\mathbf{e_j}$ continues to capture individual deviations from cluster-level means on the outcome vector $\mathbf{Y_j}$.

Controlling for classroom variation via this statistical approach allows for better estimates of student-level effects by accounting for nuances that are unique to each classroom. In this way, student effects are less confounded with classroom-specific features. In our case, we have up to six observations (2013-2018) per public protection measure for each state, and each state has a different score required to pass the bar exam. It may be argued that there are uniquenesses within a state that contribute to variation in the numbers of complaints, charges, and disciplines that could obscure the relationship between cut score and public protection. To

24

account for this, we can use multilevel modeling to account for state-specific variation. Our model inclusive of cut score as a predictor can be written as

$$Y_{ij} = \gamma_{00} + u_{0j} + \gamma_{01} \text{ cutscore}_{ij} + e_{ij},$$

Where i represents the observation (year between 2013 and 2018) and j represents the state.[69] Thus, Y_{ij} represents the ith observation for the jth state on one of the three public protection outcomes, γ_{00} is the overall average of the specified public protection outcome, u_{0j} is the state-specific adjustment to the overall average of the public protection outcome, γ_{01} is the estimated fixed effect of cut score on the public protection outcome applied to the ith observed cut score in the jth state, and e_{ij} is the residual associated with observation i in state j for the modeled public protection outcome.

Controlling for state-level variation in this way can filter out this additional noise when assessing the relationship between bar exam cut score and public protection. Our models followed suggested procedures by West et al. (2015). For a more technical look at multilevel modeling, please review West et al. (2015) or Hox & Roberts (2011).

Hypothesis Testing

Statistical models provide coefficients which can be used to test certain statistical hypotheses. Procedurally, the researcher declares a hypothesis (called the null hypothesis,

---

[69] It is assumed that within a cluster, observations are independent. In these models, the observations within the cluster correspond to states' time points. To assume independence is therefore to assume that there is no longitudinal relationship within a state's public protection measures. Recognizing that this may be a strong assumption, we constructed additional models in which time is the cluster rather than state. Unfortunately, statistical limitations of only one observation per time point for each state preclude us from controlling for both state and time random effects simultaneously. In the alternative models clustering on time, the emergent results only strengthen the findings of this paper. Consequently, we note here that these models were explored, but ultimately omitted because we felt that controlling for within-state variation offered greater theoretical interpretability than the alternative, and the choice was inconsequential to the findings of our research.

denoted H_0), then seeks evidence from a statistical model or hypothesis test to refute the declared

hypothesis. The evidence contrary to the null hypothesis is said to favor the alternative

hypothesis (denoted H_1 or H_A).

In this research, our null hypothesis is that there is no statistically significant relationship

between cut score and any of the public protection measures defined throughout this manuscript.

Stated formally, we write

$$H_0: \gamma_{01} = 0$$

$$H_1: \gamma_{01} \neq 0$$

where γ_{01} refers to the regression coefficient describing the relationship between the bar exam cut

score and the public protection measure of interest.

Any statistical evidence rejecting the null hypothesis would lead us to conclude that a

relationship may exist, and we are subsequently left to interpret its direction (positive or

negative). Positive relationships indicate that higher cut scores correspond to more complaints,

charges, and/or disciplinary actions, while negative relationships indicate that higher cut scores

correspond to fewer complaints, charges, and/or disciplinary actions. Failure to reject the null

hypothesis implies there is no evidence of a relationship between cut score and public protection.

This framing of the null and alternative hypotheses takes a neutral position on whether such a

relationship exists by seeking evidence to reject neutrality.

Outliers

As good practice, outliers in this data were filtered out on the basis of Cook's distance.[70]

Whenever Cook's distance was larger than four divided by the number of observations,[71] the

---

[70] R. Cook, *Detection of Influential Observation in Linear Regression*, 19(1) Technometrics 15 (1977).

[71] Kim Choongrak & Barry E. Storer, *Reference Values for Cook's Distance, Communications in Statistics*, 25 Simulation & Computation 691 (1996).

corresponding observation was considered an outlier and subsequently removed. For each measure, there were few outlying cases. However, the number of observations, and thus the observations deemed as outliers varied across different statistical analyses.

In order to evaluate sensitivity to outliers, we also analyzed the data inclusive of outliers. We found that there were no relevant differences with or without outliers that would change the general interpretation of the findings.[72] We therefore report only the results with outliers removed from the data, and include additional tests in supplemental materials available online.

Supplemental States

We acknowledge that there are 16 supplemental state-year observations (see Sample subsection) from which data was collected from official reports on state websites rather than the SOLD. It is understandable to suspect that this data may function in ways inconsistent with the SOLD by virtue of allowing different levels of reporting, applying alternative definitions, or through some other form of variation. As such, these observations have been flagged in the data set. All models have been run with and without these observations, including the identification of and filtering of new outliers on the basis of changes in sample size. Because this had no impact on the overall findings, we report only on the analyses inclusive of these supplemental states, and again include these additional tests in supplemental materials available online.

Annotated Disciplinary Actions

---

[72] Statistical significance varied across some models. However, the vast majority of statistically significant effects, whenever they were found, suggested that the relationship between cut score and the corresponding disciplinary outcome was positive, not negative. The few exceptions were with the specific disciplinary action private admonishment discussed elsewhere in this manuscript. This would imply that as the cut score increases, the number of complaints, charges, or disciplinary actions increases. Moreover, the effect sizes in all cases were very small. Since we sought to show that there was no negative relationship between the bar exam cut score and either complaints, charges, or disciplinary actions, we deemed these fluctuations in significance and trivial positive relationships to be not pertinent to our general conclusion.

As noted above, what is categorized as public, private, and/or disciplinary varies by state (see the SOLD for details). Accordingly, what gets factored into the total number of disciplinary actions (private, public, or collective) is affected by states' decisions. We have accounted for these nuances by analyzing the data in several ways. First, we analyze the collective total of public and private disciplinary actions taken against attorneys as determined by the ABA and produced in the SOLD report. This we take as our primary analysis of the discipline data and describe our results below. Next, we analyze every particular public and private disciplinary measure separately, ignoring the nuances detailed in states' notes, thereby treating the annotated data as equivalent to the unannotated data. This assumes that states' differences do not constitute meaningful departures from other states' reports. Lastly, we discard all annotated data and analyze only data from states which do not report annotated disciplinary outcomes. This treatment assumes that the annotated data differs from the unannotated data in an important way that might influence the findings. It also assumes that unannotated states do not differ from one another in important ways. Ultimately these decisions had no bearing on the broader results. Moreover, we take the analysis of specific disciplinary actions to be secondary to the overall research aims. Given these two things, the results section includes only the analyses of the collective total of public and private disciplinary measures provided by the SOLD. We revisit these secondary analyses in the subsequent discussion.

Analytic Tools

All analyses have been run with two statistical softwares and confirmed for agreement. The data was first analyzed using the sci-kit learn[73] and statsmodels[74] modules in Python 3.7.

---

[73] Fabian Pedregosa et al., *Scikit-learn: Machine Learning in Python*, 12 J. Machine Learning Res. 2825 (2011).

[74] Skipper Seabold and Josef Perktold, *Statsmodels: Econometric and Statistical Modeling with Python*, in Proceedings of the 9th Python in Science Conference (2010).

Next, data were confirmed for agreement with the lme4[75] and baseline statistical tools in R 4.0.2.[76] Visuals were constructed using seaborn,[77] matplotlib,[78] and ggplot2.[79] Tables were prepared using stargazer.[80]

Results

As noted, the data have been analyzed using a variety of statistical techniques and with multiple inclusion/exclusion criteria. For brevity, we only narrate the models which exclude outliers and include the supplemental states. We feel that these offer the most reliable set of analyses. As indicated throughout the results section, however, these decisions had no relevant impact on our findings. For model results from analyses including outliers and/or excluding supplemental states, please review our open science repository dedicated to these findings (to be included upon acceptance). We also include data, R code, and analyses of secondary disciplinary outcomes in accordance with transparency, open science, and best practices.[81] Statistical significance was indicated by p-values less 0.05 (conventional); strength of the relationship between cut score and public protection measures was assessed by R-squared values.

Summary of Findings

---

[75] Douglas Bates et al., *Fitting Linear Mixed-Effects Models Using lme4*, 67 J. Statistical Software 1 (2015).

[76] R Core Team, *R Foundation for Statistical Computing, R: A Language and Environment for Statistical Computing*, (2020), https://www.R-project.org/.

[77] Seaborn, https://seaborn.pydata.org/index.html (visited Feb. 13. 2021).

[78] J. D. Hunter, *Matplotlib: A 2D Graphics Environment*, 9 Computing in Science & Engineering 90 (2007).

[79] H. Wickham, ggplot2: Elegant Graphics for Data Analysis. (2016), https://ggplot2.tidyverse.org.

[80] Marek Hlavac, stargazer: Well-Formatted Regression and Summary Statistics Tables, R package version 5.2.1 (2018), https://CRAN.R-project.org/package=stargazer

[81] Maxwell Hong & Ann Moran, *An Introduction to Open Science: How to Incorporate Best Practices into Your Research*, Feb. 2019, https://www.apa.org/science/about/psa/2019/02/open-science.

Our study found no compelling evidence of a meaningful significant negative[82] relationship between states' selection of a minimum passing bar exam cut score and the number of complaints, charges, or disciplinary actions taken per 1000 attorneys. Said another way, we found no evidence that higher bar exam cut scores produce fewer complaints, charges, or disciplinary actions. These results held across all public protection outcomes, statistical modeling approaches, and decisions made pertaining to data treatment (e.g., inclusion or exclusion of outliers, supplemental states, and annotated disciplinary data). Collectively, these combined approaches searching for statistical relationships consist of over 100 statistical models.

Descriptive statistics

Descriptive statistics (e.g., mean and standard deviation) for each analytical sample vary depending on both the statistical method (e.g., linear regression or multilevel modeling) and the outcome of analysis (e.g., complaints or charges). Starting with complaints, both the linear regression and multilevel model identified the same set of outliers, and thus have equivalent fit statistics. This includes a mean of 74.15 complaints per 1000 attorneys ($SD = 32.8$) calculated from 273 observations ($N = 273$). Taking the mean across years 2013-2018, 46 states ($N = 46$) are represented in the aggregated regression analysis. The mean number of complaints per 1000 attorneys across these 46 states was 73.7 ($SD = 29.3$). For charges, the linear regression analysis

---

[82] We define "negative" on the basis of the sign (positive or negative) of the regression coefficient of a statistical model. A negative relationship would reflect that as cut scores increase, complaints, charges, and/or disciplinary actions would decrease. Statistical significance is ascribed whenever the p-value of this coefficient is < 0.05, indicating that the probability that this relationship is found by mere chance is less than 5%. We assess meaningfulness by the strength of the relationship between the cut score and outcome measure, as determined by R-squared, i.e. the percentage of variance explained in the outcome by the cut score. Across all models, the largest R-squared is .07, or 7% of variance explained. We provide the R-squared statistic whenever the regression coefficient is statistically significant, but otherwise classify 7% (or less) of variance explained as trivial and not meaningful toward identifying a relationship between cut score and public accountability.

used 270 observations ($N = 270$) with a mean number of 3.1 charges ($SD = 2.55$) per 1000 attorneys. Multilevel modeling identified many additional outliers, thus reducing the number of observations to 248 ($N = 248$). This analytic sample had a mean of 2.9 charges ($SD = 1.6$) per 1000 attorneys. Aggregating over the 2013-2018, 47 states ($N = 47$) are represented with a mean of 3.25 ($SD = 2.3$) charges per 1000 attorneys. Lastly, the analytic sample for the linear regression models on disciplinary action contained 270 observations ($N = 270$) with a mean of 4.8 disciplinary actions ($SD = 3.3$) taken per 1000 attorneys. Multilevel modeling for disciplinary action also identified many additional outliers, leaving an analytic sample size of 258 ($N = 258$). The mean of this sample was 4.32 disciplinary actions ($SD = 2.4$) taken per 1000 attorneys. After averaging across 2013-2018, 47 states ($N = 47$) were represented for the aggregated analyses. These states had a mean of 4.8 disciplinary actions taken per 1000 attorneys ($SD = 2.9$). Descriptive statistics for these analytic samples are summarized in Table 1.

Table 1. Descriptive statistics by analytic method and outcome variable

Analytic Method	Outcome Variable	Mean (per 1000 attorneys)	SD	SE	N
OLS Regression	Complaints	74.15	32.81	1.99	273
	Charges	3.10	2.55	0.16	270
	Discipline	4.79	3.27	0.20	273
Aggregated Regression	Complaints	73.69	29.33	4.32	46
	Charges	3.25	2.28	0.33	47
	Discipline	4.84	2.91	0.42	47
Multilevel Modeling	Complaints	74.15	32.81	1.99	273
	Charges	4.84	1.64	0.10	270
	Discipline	4.32	2.36	0.15	258

Regression analyses

Linear Regression

31

Beginning with the OLS regression models, we found no statistical evidence suggesting a meaningful negative relationship between states' bar exam cut score and public protection outcomes. All statistically significant relationships found were positive, though the highest R-squared value was .034 suggesting that only 3.4% of the variance in the outcome was explained. What this may mean is that, contrary to the conventional assumption that lower cut scores are associated with more complaints and discipline in a jurisdiction; the converse may be true--higher cut scores may be associated with more complaints and discipline in a jurisdiction, not less. Even so, we consider this to be a very weak relationship, if not altogether spurious, and thus choose not to draw any inferences from or build discussion around these findings.

Specific findings were as follows. The relationship between states' cut score and the number of complaints filed per 1000 attorneys was significant and positive, but weak ($b = 1.95$, $p < .01$, $R^2 = .037$). The relationship between cut score and the number of charges filed after probable cause per 1000 attorneys was similarly positive, statistically significant and weak ($b = 0.15$, $p < .01$, $R^2 = 0.035$). Both of these relationships hold whenever supplemental states are excluded. Including outliers eliminates all significant relationships, with or without supplemental states. OLS regression found no relationship between the number of disciplinary actions filed against attorneys and states' bar exam cut scores ($b = 0.00$, $p = 0.96$, $R^2 = 0.00$). This held true independent of inclusion or omission of outliers, and/or inclusion or omission of supplemental states. Results for OLS regression models are summarized below in Table 2. We also visualize the relationship between bar exam cut scores, complaints per 1000 attorneys, charges per 1000 attorneys, and discipline per 1000 attorneys in Figures 1-3, respectively.

Table 2. Results of OLS regression for public protection outcomes

Dependent Variable

32

	Complaints per 1000 (SE)	Charges per 1000 (SE)	Discipline per 1000 (SE)
Cut Score	1.95*** (0.61)	0.15*** (0.05)	-0.0004 (0.06)
Constant	-189.04** (81.52)	-16.83*** (6.35)	4.84 (8.28)
Observations	273	270	273
R^2	0.037	0.035	0.000
Adjusted R^2	0.034	0.032	-0.004
Residual Std. Error	32.25 (df = 271)	2.51 (df = 268)	3.27 (df = 271)
F-Statistic	10.43*** (df = 1; 271)	9.85*** (df = 1; 268)	0.000 (df = 1; 271)

Note: *p < 0.1; **p < 0.05; ***p < 0.01

Figure 1. Complaints per 1000 Lawyers by Cut Score 2013-2018.

Complaints Per 1000 Lawyers by Cut Score 2013-2018

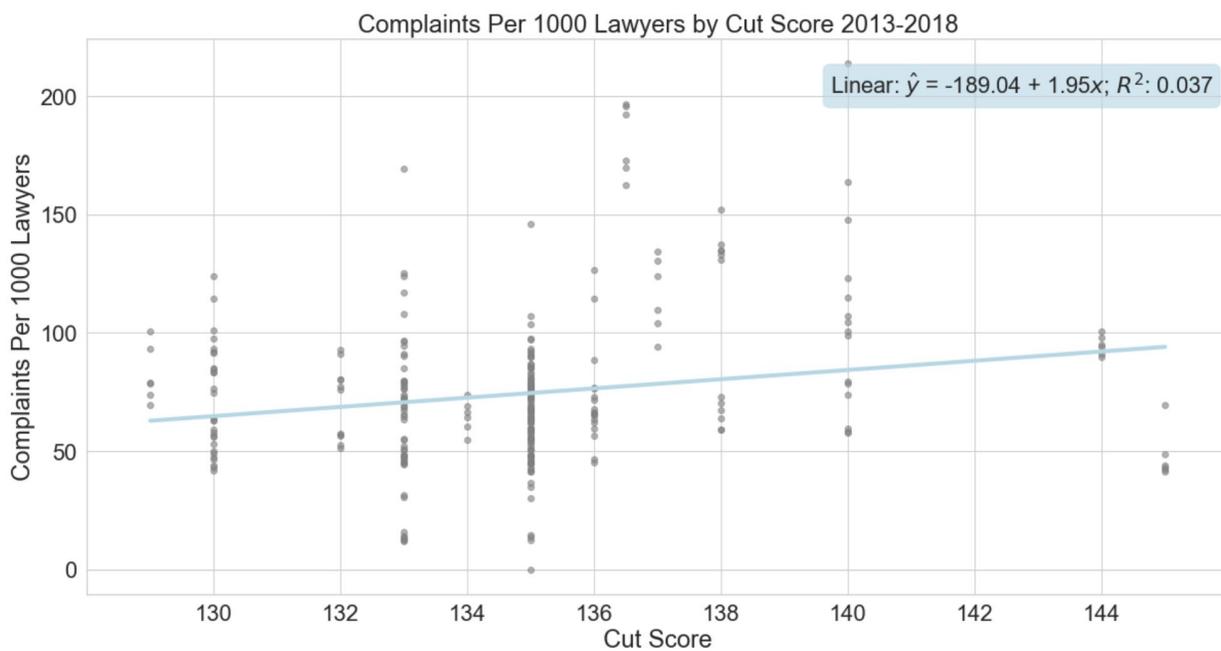Linear: $\hat{y}$ = -189.04 + 1.95x; R^2: 0.037

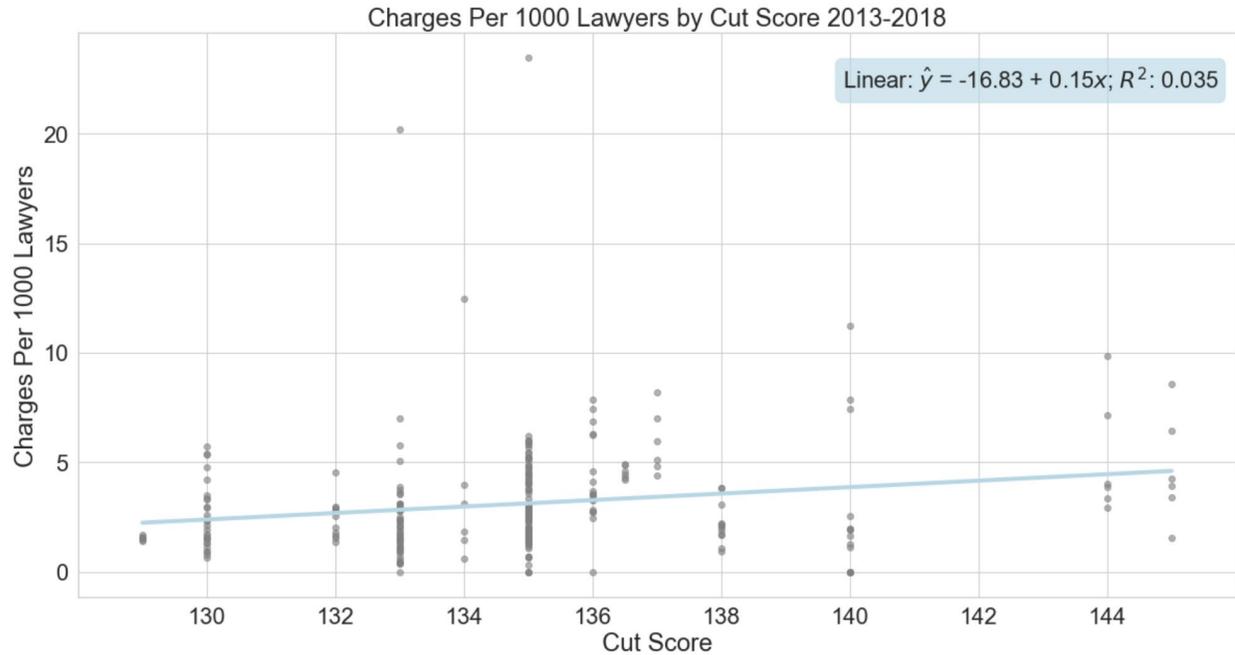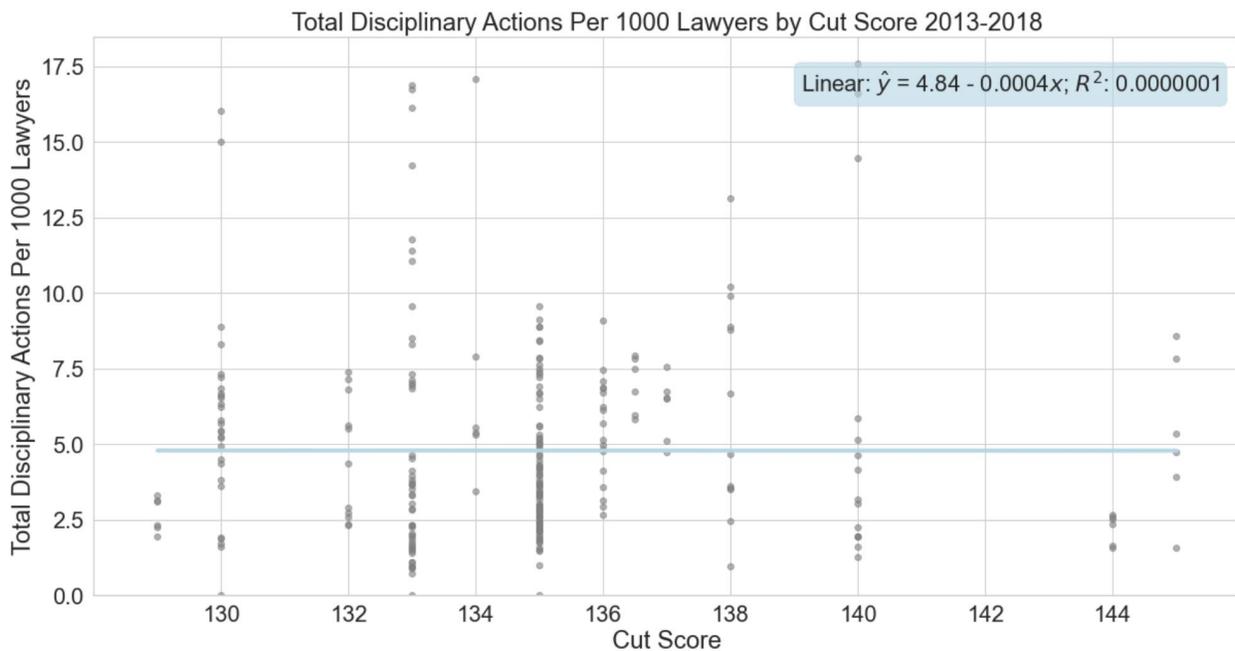Figure 2. Charges per 1000 Lawyers by Cut Score 2013-2018.

Figure 3. Total Disciplinary Actions per 1000 Lawyers by Cut Score 2013-2018.

Aggregated Linear Regression

Turning to the OLS models in which each state's average across all years was analyzed, we again found no meaningfully significant negative relationships between states' minimum passing bar exam cut score and public protection measures. Statistically significant relationships

34

were fewer than those found in the disaggregated OLS models, and in fact were only marginally significant. These relationships were once again positive and weak wherever present. The largest R^2 for these models was 0.065.

Specific findings were as follows. We observed a marginally significant and positive relationship between states' bar exam cut scores and the number of complaints filed per 1000 attorneys ($b = 2.63$, $p = .09$, $R^2 = .065$). The relationship between cut score and charges filed per 1000 attorneys was non-significant ($b = .078$, $p = .46$, $R^2 = .01$), as was the relationship between cut score and the number of disciplinary actions taken per 1000 attorneys ($b = -.06$, $p = .70$, $R^2 = .003$). These results hold independent of whether supplemental states are included or excluded, and independent of whether outliers are retained or omitted. Aggregated regression results are summarized below in Table 3.

Table 3. Results of Aggregated OLS regression for public protection outcomes

	Dependent Variable		
	Complaints per 1000 (SE)	Charges per 1000 (SE)	Discipline per 1000 (SE)
Cut Score	2.63* (1.51)	0.08 (0.11)	-0.06 (0.15)
Constant	-280.16 (202.38)	-7.32 (14.26)	12.76 (20.17)
Observations	46	47	47
R^2	0.065	0.012	0.003
Adjusted R^2	0.044	-0.01	-0.02
Residual Std. Error	28.68 (df = 44)	2.29 (df = 45)	2.94 (df = 45)
F-Statistic	3.06* (df = 1; 44)	0.55 (df = 1; 45)	0.15 (df = 1; 45)
Note:			*$p < 0.1$; **$p < 0.05$; ***$p < 0.01$

Multilevel Modeling

Results of multilevel regression analyses were similar to those of the aggregated and disaggregated OLS models. We find no meaningfully significant negative relationships between

states' minimum passing bar exam cut score and our public protection measures. All but one regression coefficient are positive, and none of them are statistically significant at $p < .05$. Unlike OLS regression models, however, multilevel modeling does not provide R^2 statistics. To supplement this, we use Akaike's Information Criteria (AIC), Bayesian Information Criteria (BIC), and the Likelihood Ratio Test (LRT) to describe the fit of the model with and without cut score as a predictor. AIC and BIC are designed to favor parsimony by precluding superfluous variables from entering the model. For both statistics, lower values reflect a better model; if two models have approximately similar AIC and BIC, the more parsimonious model is recommended. The LRT tests for a statistically significant difference between two models. If the test is not statistically significant, there is no reason to believe a difference exists between the two models and thus the more parsimonious of the two is preferred. When a statistically significant difference is present, the model with the higher log-likelihood value is preferred.

Specific findings were as follows. There were no statistically significant relationships between states' bar exam cut scores and the number of complaints filed per 1000 attorneys ($b = 2.41$, $p = 0.1$). AIC (2363.8) and BIC (2378.2) for the model containing cut score as a predictor of complaints were only marginally different from those without cut score as a predictor (AIC = 2364.7; BIC = 2375.6). The LRT suggests there is no statistically significant difference between the models with or without cut score as a predictor ($\chi^2(1) = 2.91$, $p = .09$). Both AIC/BIC and the LRT suggest that the cut score does not contribute to predicting the number of complaints per 1000 attorneys. Similarly, there was no statistically significant relationship between states' bar exam cut score and either the number of charges filed after probable cause per 1000 attorneys ($b = .101$, $p = 0.13$). AIC (784.5) and BIC (798.6) for the model including cut score were notably larger than the model omitting cut score (AIC = 735.1; BIC = 745.6), suggesting the inclusion of

36

cut score is superfluous. Additionally the LRT suggested a strong difference between the model

with or without cut scores included as a predictor, but in favor of the more parsimonious model

in which cut score was omitted ($\chi^2(1) = 47.4$, $p < .001$). Thus, the multilevel models suggest there

is no evidence to support a meaningful relationship between cut score and the number of charges

filed per 1000 attorneys. Lastly, we found no relationship between cut score and the number of

disciplinary actions taken per 1000 attorneys ($b = -.02$, p = .87). The AIC, BIC, and LRT

findings were similar to those of the charges models. AIC (964.2) and BIC (978.4) for the model

with cut score as a predictor were much larger than those in the model that did not include cut

score (AIC = 878.6; BIC = 889.1). Similarly, the LRT suggested a strong statistical difference

between the models, again favoring the model omitting cut score ($\chi^2(1) = 83.6$, $p < .001$). Taken

together, these findings suggest there is no evidence of a relationship between the bar exam cut

score and disciplinary actions taken per 1000 attorneys. As with other models, these results held

independent of whether supplemental states were included or excluded, and independent of

whether outliers were retained or omitted. Results for multilevel modeling are summarized

below in Table 4.

Table 4. Results of multilevel modeling for public protection outcomes

	Dependent Variable		
	Complaints per 1000 (SE)	Charges per 1000 (SE)	Discipline per 1000 (SE)
Cut Score	2.41* (1.42)	0.10 (0.07)	-0.02 (0.11)
Constant	-249.62 (191.29)	-10.67 (8.82)	6.91 (14.20)
Observations	273	248	258
Log Likelihood	-1174.23	-388.25	-478.0
AIC	2356.46	784.50	964.21
BIC	2370.89	798.56	978.42

High Level Lag Analysis

Prior research has argued that lawyer discipline seldom occurs during the first years of practice. A lag is posited to exist between the time a lawyer passes the bar and when discipline occurs. Anticipating this lag effect, we conducted additional analyses controlling for its impact. These models utilize states' cut scores from 2005 rather than their current cut scores. Given that our discipline data ranges from 2013 to 2018, this gives us an 8- to 13-year lag window.

Linear regressions, aggregated regressions, and multilevel models were again used to evaluate the impact of lag analyses. Results of these analyses were narratively indistinguishable from analyses using current cut scores. That is, we found no significant negative relationship between cut scores and disciplinary actions against attorneys. Moreover, the number of disciplinary actions taken against attorneys per 1000 continues to be extraordinarily small. Given that the multilevel models offer the most rigorous statistical approach and that there were no narrative changes across other models, we report only the multilevel model results here (see Table 5). AIC, BIC, and log-likelihood statistics revealed no significant changes to the models with and without cut score included as a predictor, thus suggesting it is not a useful predictor of lawyer discipline. Collectively, these lag analyses suggest prima facie evidence[83] that lawyer discipline is not driven by older attorneys.

---

[83] The small number of disciplinary actions taken per 1000 attorneys suggests that additional analyses of lag effects would be of limited use from a policy perspective, though it could be methodologically interesting. Additional models using cut scores from different time windows (e.g., 15 years, 20 years, 30 years, etc.) could be analyzed. Alternatively, statistical models accounting for the amount of time an attorney has practiced law could be instrumental both in

Table 5. Lag analysis results of multilevel modeling for public protection outcomes

	Dependent Variable		
	Complaints per 1000 (SE)	Charges per 1000 (SE)	Discipline per 1000 (SE)
Cut Score	0.78 (1.10)	0.09 (0.05)	-0.07 (0.08)
Constant	-31.28 (147.18)	-8.80 (6.57)	14.37 (10.60)
Observations	260	235	245
Log Likelihood	-1121.43	-371.35	-455.78
AIC	2250.86	750.71	919.57
BIC	2265.10	764.55	933.57

Note: $*p < 0.1; **p < 0.05; ***p < 0.01$

Discussion

The choice of a cut score on the bar exam is also a choice about the size and diversity of the profession. Lower cut scores would provide the public greater access to a more representative set of lawyers. But perhaps costs exist that some believe would outweigh these benefits. Certainly, existing cut scores are often defended as instruments for public protection. Our analyses explore whether lower cut scores would result in higher rates of discipline per lawyer on the basis of the metrics chosen by legal regulators and the ABA. We find no evidence to support such an assertion.

Descriptive & Statistical Findings

Three main takeaways emerge from our analyses. First, results suggest that the mean number of complaints is approximately 74 per 1000 attorneys (see Table 1). For charges, the mean shrinks to around 4 per 1000 attorneys, and roughly 4.6 per 1000 attorneys for discipline.

---

identifying when (if ever) in an attorney's career they are most likely to be disciplined, and in more precisely controlling for the lag effect.

While variability between states is expected, these averages suggest that the overall number of attorneys with complaints, charges, and disciplinary actions taken against them is quite small in relation to the overall population of attorneys. Said another way, between 2013 and 2018 less than 10% of attorneys had complaints filed against them (7.4% to be exact), and less than half a percent faced charges or disciplinary action (0.4% and 0.46% respectively). With so few cases, even the best pre-practice predictors of subsequent discipline would, at best, be weak.[84] Second, statistical models consistently fail to find a meaningful statistically significant negative relationship between cut score and public protection measures. Across the majority of our statistical models, we fail to reject the null hypothesis that the relationship between cut score and public protection is nil. Third, whenever these relationships are significant (though weak and unmeaningful), they are almost always in the positive direction, which would imply heightened cut scores correspond with more complaints, charges, and/or discipline. The most robust of these findings--those from multilevel modeling--suggest no significant relationship at all (but nevertheless a positive trendline).

Taken together, the findings fail to support the claim that higher bar exam cut scores correspond to greater public protection. If anything, statistical evidence points in the opposite direction: higher bar exam cut scores may lead to less public protection. These general findings remain consistent even across numerous statistical modeling approaches, controlling for state-level peculiarities via multilevel modeling, modifying inclusion/exclusion criteria, and as discussed below, analyses of specific disciplinary actions. Not only is there no evidence of a significant negative relationship between cut score and public protection, this empirical study reveals that the phenomenon, at issue, involves a small minority of the profession who

---

[84] *See* LEVIN ET AL., *supra* note 32 (so finding).

experience public or private discipline to begin with. This minority is by all statistical reasoning unaffected by the choice of bar exam cut score, and evidently unaffected in the way proponents of higher cut scores contend. Consequently, we reject on the basis of no supporting evidence the argument that heightened bar exam cut scores increase public protection. Indeed, their empirically demonstrated effect is to reduce diversity and inclusion within the legal profession with no apparent corresponding benefit. The policy prescription is worse than the alleged disease--and ineffective against it.

Secondary Disciplinary Analyses

As indicated in the methods section, we also explored the relationship between cut scores and specific disciplinary actions. These included private admonition, private reprimand, letters of warning, involuntary disbarment, disbarment on consent, suspension (excluding interim suspension), interim suspension (risk of harm or criminal conviction), public admonishment/reprimand/censure, probation, order to pay restitution, and order to pay costs. We omitted these models from the results largely for the sake of brevity. Variations of inclusion/exclusion criteria and modeling approaches for these specific public protection outcomes yield nearly 200 statistical models and findings and would thus be too much to individually report on. Moreover, the findings fail to alter the conclusion.

At a high level, none of these findings change the overall narrative. As we inflate the number of models, we would expect 5% of models to be statistically significant by mere chance. And indeed, we do find more statistically significant results, but they continue to be weak and most frequently in a *positive* direction that suggests higher cut scores lead to more disciplinary actions taken. However, some are negative. In particular, a weak statistically significant pattern emerges suggesting that higher cut scores result in fewer *private* admonitions. Yet, this pattern is

counterbalanced with other emergent results suggesting that higher cut scores also result in more attorneys placed on probation, more attorneys disbarred on consent, and more attorneys suspended. In all cases, correlations between cut scores and outcomes are weak. Moreover, these emergent relationships largely vanish when state peculiarities are accounted for by using multilevel models. Given the weakness and ephemerality of these findings, we therefore find no convincing evidence of a meaningful negative (or positive) relationship between public protection and bar exam cut score.

Consequential Validity

The lack of statistical evidence that higher bar exam cut scores improve public protection suggests that they lack consequential validity (i.e., that their positive social consequences do not outweigh their negative ones). Consider the entries on the other side of the ledger. The bar exam is a Jim Crow relic that somehow survived the Civil Rights era despite failing to satisfy substantive Title VII antidiscrimination standards. It was forged to achieve the exclusion that still defines it and that high cut scores exacerbate. Racial and ethnic minorities remain grossly underrepresented in the profession to the mutual detriment of the public and the profession. These attorneys are more likely than their White peers to start their careers in government service, public service, or public interest work. They also provide service to minority clients, engage in pro bono work, sit on community organizations boards, and mentor younger attorneys at higher rates. As a result, the harms of excluding such attorneys from practice fall disproportionately on potential clients who have limited means or hail from underrepresented racial groups.

Arbitrarily high cut scores also entrench the justice gap, especially for those with the fewest resources. Consider the largest U.S. jurisdiction, California, where a recent study found

42

that 85% of members of the public with civil legal problems failed to receive adequate legal help. The problem was worse for those living in poverty, but even among Californians earning more than 600% of the federal poverty line, more than three quarters lacked adequate legal help.[85] Had California joined Minnesota in setting its cut score at 1300 rather than 1440, it would have licensed 12,907 more attorneys between February 2009 and July 2018, including more than 6,000 new attorneys who are people of color, a 26.4% uptick in the number who entered the profession.[86] Moreover, these lawyers would have been particularly likely to provide services to the people who most needed them.

Policy Recommendations for Meaningful Public Protection

It is time for legal regulators to look to new solutions. Research offers abundant alternatives to the failed effort to weed out "bad apples" with a test not designed to do so. Behavioral legal ethics and evidence-based reforms are particularly promising pathways to pursue. Most obviously among potential reforms, cut scores could be lowered. Law school graduates who fall somewhat short of passing the bar could be permitted to enter practice following an apprenticeship--a variation on a once-common practice. Post-entry training could also be improved. Medicine requires residencies, many foreign jurisdictions have multi-month bridge-the-gap programs, and continuing legal education in one's practice area could be

---

[85] State Bar of California, California Justice Gap Study: The Service Gap -- Findings and Recommendations (2020), 1, at http://www.calbar.ca.gov/Portals/0/documents/accessJustice/Justice-Gap-Fact-Sheet-Service-Gap.pdf.

[86] 46.9% of newly admitted attorneys would have been attorneys of Color, plus another 392 examinees whose racial/ethnic group is unknown. *See* Winick et al., *supra* note 4, at 27. This report does not include the number of non-responses or the number of participants who responded "other" as their racial/ethnic group. However, the authors have access to the Winick et al. data and have used it to update the numbers reported here. Responses of "other" are included as people of color, and non-responses were added to the denominator but treated as neither White nor people of color.

mandatory.[87] To address public defender systems and large firms outperforming assigned defense counsel and smaller and solo firms, one could require robust lawyer assistance programs, greater on-the-job mentoring requirements (from within or outside a firm or office), recordkeeping support or administrative oversight for solo and small firms, and mandatory public defender systems.[88] Measures of competence could also be improved, including by recognizing that the best time to prevent incompetent practice is during practice. Currently, there is a single ill-suited test of lawyerly competence before practice and primarily piecemeal, reactive, self-dealing, client-initiated, and often under-resourced disciplinary proceedings subsequently.[89] But jurisdictions could require a series of tests of increasing difficulty, testing of active lawyers on their practice areas,[90] mandatory structured peer review that could include detailed and area-specific process-based checklists, use in evaluation of dummy cases, and better-resourced disciplinary bodies with more substantive non-lawyer participation.[91]

---

[87] *See* Krivosha, *supra* note 31; Martyn, *supra* note 31, at 729; American Bar Association Section of Legal Education and Admissions to the Bar, Legal Education and Professional Development-An Educational Continuum (1992); Jayne W. Barnard and Mark Greenspan, *Incremental Bar Admission: Lessons from the Medical Profession*, 53 J. Legal Educ. 340J (2003).

[88] Jeff Giddings, *Legal Aid Services, Quality and Competence: Is Near Enough Good Enough and How Can We Tell What's What?* 1 Newc L.R. 66 (1996); Cranton, *supra* note 31, at 6; Merritt, *supra* note 33; Barnard et al., *supra* note 87.

[89] Krivosha, *supra* note 31, at 828-829; Martyn, *supra* note 31; Cranton, *supra* note 31, at 10. An alternative to disciplinary hearings is civil malpractice suits. These also begin with a client's action, but are adjudicated by judges and juries rather than lawyers. Such actions are limited, however, by their heavy focus on lack of legal knowledge, which is but one form of incompetence. *See* Martyn, *supra* note 31, at 734; *cf.* Cranton, *supra* note 31, at 10 (suggested expanded malpractice liability).

[90] Barnard et al., *supra* note 87; Wegner, *supra* note 31; Krivosha, *supra* note 31, at 828-829; Cranton, *supra* note 31, at 10.

[91] Giddings, *supra* note 88; Martyn, *supra* note 31, at 729; Cooper, *supra* note 28, at 115-117; Cranton, *supra* note 31. One alternative that some jurisdictions have taken up is to create alternative forums (e.g. fee dispute resolution systems). *See* Martyn, *supra* note 31, at 731. Here, the stakes are lower than in discipline proceedings, which may make attorney adjudicators more willing to articulate standards of competence. *See id.* Output measures are also available. Those such as award size or jail time are context specific. Cooper, *supra* note 28, at 116-117. Consumer

A Failed Approach

It is no longer tenable to judge the public protection benefits of higher cut scores to outweigh their myriad costs. There is no good evidence demonstrating that bar exam scores and subsequent discipline are related. The profession also attaches low value to public protection, as evidenced by its reluctance to implement measures that promote lawyer competence and by the infrequency with which it metes out discipline. Were legal regulators to decide that lawyer incompetence is more common or more serious than their actions currently suggest, policy responses other than high cut scores would be more effective. Such responses would also incur fewer social costs than higher cut scores, which produce inequality, discrimination and underrepresentation. Worse, high cut scores violate notions of just deserts by punishing people for bad propensities rather than bad acts and by denying them opportunities earned through toil, sacrifice, investment, and learning. The problems are interconnected. Social realities of inequality, underrepresentation, and discrimination too often overawe efforts to overcome needlessly high cut scores for candidates who have matriculated from colleges and universities and successfully completed law schools, demonstrating both mastery and repeated successes throughout their careers.

Conclusion

It has been more than half a century since Congress enacted substantive racial antidiscrimination standards for employment tests that the bar exam does not meet. So reasoned a United States Court of Appeals in 1972. This moral failure is not redeemed by the legal technicalities that have nonetheless permitted the bar exam to continue. The legal profession

---

satisfaction has the problem that much representation iis invisible to the client or beyond her competence. *Id*.

violates its ideals of justice and fairness and harms the public by screening out qualified aspiring lawyers who are disproportionately from underrepresented and disadvantaged groups.

It is no answer that high cut scores prevent attorney malfeasance. That claim has always been dubious on its face. The bar exam was not designed for that purpose, and most lawyer discipline comes decades after the administration of the test. Yet the question was difficult to study empirically given limitations of the data. Indeed, the main consequence of prior studies was undue public endorsement of an unlikely hypothesis. This study overcomes some of the data's limitations to confirm common sense. After rigorous statistical investigation, we find no evidence to support the claim that lower cut scores lead to greater lawyer misconduct. The discussion should thus be shifted from unsupported claims of malfeasance to the demonstrable public benefits granted by lower cut scores. Heightened cut scores are professional gatekeeping masked as public protection.

Absent new research providing strong, heretofore overlooked evidence of a meaningful negative relationship between bar exam cut score and public protection, the time has come for legal regulators to undertake the empirically validated path to public protection that does exist: lowering cut scores. Doing so would meet twin legal crises: lack of diversity in the profession and lack of access to justice for all. It would grow the bar, diversify practice, drive upward mobility, and add lawyers more likely to take on government, community, public-interest, and underrepresented-minority representations. In this way, the demonstrable tragedy of high cut scores point the way to the ready-at-hand opportunity of less-punishing replacements.