

Putting the Bar to the Test: An Examination of the Predictive Validity of Bar Exam Outcomes on Lawyering Effectiveness

Journal of Law and Empirical Analysis
July-December 2024: 1–13
© The Author(s) 2024
Article reuse guidelines:
sagepub.com/journals-permissions
DOI: 10.1177/2755323X241296683
journals.sagepub.com/home/lex



Jason M. Scott¹, Stephen N. Goggin² , Rick Trachok³, Jenny S. Kwon⁴, Sara Gordon⁵, Dean Gould⁶, Fletcher S. Hiigel¹, Leah Chan Grinvald⁷, and David Faigman⁴

Abstract

How well does bar exam performance predict lawyering effectiveness? Is performance on some components of the bar exam more predictive? The current study, the first of its kind to measure the relationship between bar exam scores and a new lawyer's effectiveness, evaluates these questions by combining three unique datasets—bar results from the State Bar of Nevada, a survey of recently admitted lawyers, and a survey of supervisors, peers, and judges who were asked to evaluate the effectiveness of recently-admitted lawyers. We find that performance on both the Multistate Bar Examination (MBE) and essay components of the Nevada Bar have little relationship with the assessed lawyering effectiveness of new lawyers, calling into question the usefulness of these tests.

Keywords

bar exam, MBE, Nevada bar, lawyering effectiveness, validity

Introduction

Virtually all jurisdictions in the United States require law school graduates to pass a bar examination before being licensed to practice.¹ According to the National Conference of Bar Examiners (NCBE), which develops and administers the Uniform Bar Examination (UBE) and the Multistate Bar Examination (MBE), the purpose of these exams is to ensure that only individuals who possess minimum competence to practice law are granted licenses to do so ([National Conference of Bar Examiners \[NCBE\], 2021](#)). Bar exams are thus broadly understood as protecting the public from incompetent lawyers. Given this purpose, the primary utility of bar exams should depend on whether exam scores distinguish the competent from the incompetent. Yet, for many people who have taken a bar exam and then practiced law, the relationship between scores and competence appears weak.

Bar exams principally test memorization of legal doctrine, procedural knowledge, analytical reasoning, essay writing, and time management. Although these skills are related to law

practice, they are but a small part of what it means to be a competent attorney.

The bar exam also involves, to some degree, assessing test-taking abilities, but these are not requisite skills for being a competent attorney. Although test-taking strategies can be

¹AccessLex Institute, West Chester, PA, USA

²Department of Political Science, San Diego State University, San Diego, CA, USA

³Nevada Board of Bar Examiners, Reno, NV, USA

⁴University of California College of the Law, San Francisco, San Francisco, CA, USA

⁵Peter A. Allard School of Law, University of British Columbia, Vancouver, BC, Canada

⁶State Bar of Nevada, Reno, NV, USA

⁷William S. Boyd School of Law, University of Nevada, Las Vegas, Las Vegas, NV, USA

Corresponding Author:

Stephen N. Goggin, Department of Political Science, San Diego State University, 5500 Campanile Drive, San Diego, CA 92182, USA.
Email: sgoggin@sdsu.edu



(and often are) taught in law schools, some individuals seem to be innately better at taking tests—no matter whether this can be attributed to genetic or environmental factors, or some combination of the two (see, e.g., [Plomin, 2023](#); [Plomin et al., 2014](#); [Wadsworth et al., 2002](#)).

Despite their widespread use, the key question of whether bar exams are useful indicators of minimum competence—and by extension, the extent to which lawyers can effectively perform their duties—deserves careful scrutiny, something to which it has rarely been subjected.

Adding to the need for validity testing are the persistent racial disparities in bar passage rates and representation in the legal profession. According to American Bar Association (ABA) data, in 2021, White law school graduates had first-time pass rates 24 percentage points greater than their Black peers, 13 percentage points greater than their Hispanic peers, and 15 percentage points greater than their Native American peers. These disparities translate to representation in the legal profession. Data from the ABA National Lawyer Population Survey, which provides the most comprehensive picture of the demographic composition of the legal profession in the U.S., show that, in 2022, only 19% of practicing attorneys nationwide were people of color. By comparison, people of color comprised 40% of the U.S. population.

An exam, such as the bar exam, that restricts access to the profession in such a disparate manner should be subject to frequent, public, and rigorous validity testing.

This study, a multi-institutional collaboration, directly assesses the extent to which performance on the bar exam (in its prevailing current format) is associated with the skills and knowledge lawyers use on a day-to-day basis. It is the first such study to use a representative statewide sample to explore these relationships and ours is the first researcher team to be granted permission by Professors Marjorie Shultz and Sheldon Zedeck to use their innovative instruments for measuring lawyering skills—developed over decades of researching this issue.

Using Tobit regression to analyze survey data obtained from more than 500 recently licensed attorneys in Nevada, we explore the extent to which these attorneys' performance on the bar exam explains ratings they receive from peers, supervisors, and judges on their lawyering effectiveness. Overall, we find that although in some cases positive relationships exist, the effects are small and lack practical significance.

Although the bar exam is slated for a significant overhaul in the form of the NextGen Bar Examination, it will retain similar components, namely the multiple choice and practice essay components. The need to memorize and work within a tightly-timed environment will also remain inherent. Thus, the results from this study, although specifically focused on the current form of the bar exam, should remain applicable when the NextGen Bar Exam is released in 2026. Moreover, our work may help jurisdictions inform decisions regarding cut scores, particularly given our findings that an attorney's

performance on the bar exam has little relationship with ratings of their lawyering effectiveness. In light of these findings, jurisdictions may be less reluctant to reduce their cut scores or consider additional licensure pathways, something several jurisdictions are currently considering or have recently adopted.

The Bar Exam

Although its specific composition varies by jurisdiction, in general, the bar exam comprises three parts: a multiple-choice component, a subject-specific essay component, and a performance test component. Unlike the first two, the latter is not a test of substantive knowledge; rather it tests the ability to apply certain skills that lawyers are expected to demonstrate.

Most examinees take the Uniform Bar Examination (UBE), which is administered in 41 jurisdictions (including the District of Columbia and Virgin Islands). The UBE comprises: the Multistate Bar Examination (MBE), a six-hour, 200-question multiple-choice exam split into two sessions with no scheduled breaks during either session; the Multistate Essay Examination (MEE), composed of six 30-min essay questions; and the Multistate Performance Test (MPT), consisting of two 90-min tasks. Some jurisdictions include additional, state-specific components. Of the non-UBE jurisdictions, all but Louisiana administer at least the MBE, and six jurisdictions administer all three components in addition to their own exams, though this is set to change in California and Nevada, which are proceeding with plans to administer new exams that comprise no components developed by the NCBE.

Among UBE jurisdictions the scores on the MEE and MPT are scaled to the MBE and total scores are calculated by the NCBE. In calculating the total score, the MBE constitutes 50% of the final score, the MEE 30%, and the MPT 20%.

Regardless of where the examinee sits for the bar exam, the exam's purpose is to distinguish those who "possess the minimum knowledge and skills to perform activities typically required of an entry level lawyer" ([NCBE, 2021](#), p. 2). Those who meet or exceed a particular score are considered "minimally competent." What score constitutes this threshold for competence varies by jurisdiction. Surprisingly, even among those 41 jurisdictions that administer the UBE, which is uniformly administered and scored (by the NCBE), the cut score ranges from 260 to 270 points out of 400 points (until February 2023 the upper bound of the range was 280 points).

These cut scores are determined as a matter of policy within jurisdictions; they are most typically reflections of values rather than the products of technically rigorous methods.

This means that cut scores are subject to considerable discussion and debate within the legal profession and the legal academy (e.g., [Winick et al., 2020a, 2020b](#)). Some commentators argue that cut scores are set too high, resulting in minimally competent bar candidates being excluded from

the practice (Rosin, 2008). Others argue that cut scores should be maintained at current levels (Anderson & Muller, 2019; Kinsler, 2017). Another thread of commentary regards the variability in cut scores, with some critics asserting that a nationally uniform score is needed, while others see the diversity of cut scores as reflections of jurisdiction-specific needs and contexts (Howarth, 2017; Merritt et al., 2001; Rosin, 2008).

These arguments, however, rest upon an assumption that if a particular score can distinguish minimum competence, then from a psychometric standpoint, the test itself must be assessing lawyering competence. However, there are reasons to doubt that this is the case—and it would be cause for concern if minimum competence were assessed using an instrument that actually measures some other construct. Therefore, one should expect that as bar exam scores increase, so too does “competence.” Thus, for bar exam scores to be validly interpreted as indicating minimum competence, the exam must first be demonstrated to be actually assessing lawyering competence.

This requires that “competence” be defined and that a system of measuring it be both devised and validated. Such a definition has not been provided by the NCBE; however, beginning in the 1970s researchers began cataloging the competencies necessary for effective lawyering, primarily by surveying practicing lawyers about the skills they valued as important for their work (Benthall-Nietzel, 1974; Schwartz, 1973).

Most recently, an effort led by Professor Deborah Merritt and the Institute for the Advancement of the American Legal System developed a set of 12 “building blocks” of minimum competence and recommendations for assessing each. These 12 competencies were drawn from qualitative work, across 50 focus groups composed of a diverse collective of more than 200 participants.

The 12 “building blocks” share many similarities with the 26 “effectiveness factors” identified by professors Shultz and Zedeck (2011). Shultz and Zedeck’s work was pioneering in that it set out to not only identify the competencies lawyers need to practice law but to create an assessment instrument using a behaviorally anchored rating scale (BARS) methodology.

Beginning with interviews and focus groups with faculty, students, and alumni of UC Berkeley School of Law, Shultz and Zedeck identified 26 lawyering effectiveness factors (which they grouped into eight thematic categories), along with hundreds of behavioral examples to illustrate how a lawyer would demonstrate various levels of performance within each factor. Shultz and Zedeck then asked 9555 law school alumni to review the examples and indicate the extent to which they aligned with each performance level of each skill, finding general agreement between the accuracy of the examples and the levels of lawyering effectiveness they represented. These 26 factors were then validated through a series of surveys of 1148 law school alumni and ratings of

those alumni that were provided by supervisors and peers (Shultz & Zedeck, 2011). Several of these noncognitive skills are also found in earlier work defining the competencies necessary to be effective lawyers, such as the MacCrate Report (American Bar Association [ABA], 1992).

Surprisingly little research exists on the topic of whether the bar exam actually assesses competence to practice law. Foster (2021) conducted the only study of which we are aware that empirically tests whether MBE scores are related to lawyering effectiveness. Foster recruited 16 practicing attorneys, all identified as “super lawyers,” to take a simulation MBE. He finds that none of the 16 participants would have achieved a passing score and that those who had practiced the longest answered fewer questions correctly (Foster, 2021). There are limitations to this approach, namely that the participants had no incentive to maximize their scores—admittance to the bar was not at stake—and the findings may not be generalizable given the size and representation of the participants, but the results are sobering and imply the need for greater scrutiny of the relationship between bar exam scores and lawyer competency.

Most of the remaining extant literature on the topic has focused largely on using occurrences and rates of attorney discipline (e.g., suspension and disbarment) as measures of lawyering competence. Kinsler, for example, studied disciplinary data for 7256 practicing attorneys in Tennessee and found that lawyers who passed the bar exam on their second attempt were twice as likely to be disciplined as those who passed on their first (Kinsler, 2017). Anderson and Muller (2019) argue that there is a relationship between bar passage scores and attorney discipline. They use logistic regression to model the effect of bar exam scores and years since admission to the bar on predicted attorney discipline—which they model as a dichotomous variable (an attorney is either disciplined or not). Anderson and Muller conclude that lowering the California MBE cut score would lead to a noticeable increase in rates of attorney discipline and consequently would present a danger to the public in the form of weakened consumer protection.

There are significant limitations to the use of lawyer discipline as proxy for lawyer effectiveness (or ineffectiveness), however. Firstly, rates of lawyer discipline are exceedingly low. In 2019, approximately one-quarter of 1% of all barred attorneys in the United States were disciplined for unethical conduct—and only one fifth of those who were disciplined were ultimately disbarred (ABA, 2022). Thus, using lawyer discipline as a measure of effectiveness yields only the blunt conclusion that virtually all lawyers are sufficiently competent (or lucky) enough to not face professional sanctions. But surely there are varying levels of effectiveness even among undisciplined lawyers. This level of granularity is not available in lawyer discipline data. A lawyer is either publicly disciplined or is not. Moreover, not all forms of misconduct are detected at similar rates. Patton’s critique of Kinsler’s and Anderson and Muller’s studies notes that

prosecutorial misconduct, for example, is rarely sanctioned. And since discipline typically follows from complaints, it would be a mistake to suppose that rates of attorney discipline reflect a systematic survey of attorney discipline (Patton, 2020).

Secondly, using whether a lawyer is disciplined requires that equal weight be assigned to each instance of discipline, regardless of the complaint or its relevance to the type of knowledge or skill tested on bar exams. Consider two hypothetical lawyers who are publicly disciplined for their conduct: Lawyer A is disciplined for failure to keep multiple clients informed about their cases and Lawyer B is disciplined for negligent failure to preserve client property. According to prevailing research frameworks, both lawyers are ineffective (or incompetent) to some degree. The violation committed by Lawyer A, however, may more accurately reflect poor communication skills (a skill not tested on the bar exam) rather than a lack of legal knowledge.² Thus, the theoretical link between the exam and the outcome is tenuous. Moreover, focusing on the *types* of discipline imposed on different lawyers adds no useful nuance. If both hypothetical lawyers are suspended from the practice, prevailing research frameworks will once again treat them the same, even though only Lawyer B has been found to lack legal knowledge.

Thirdly, rates of lawyer discipline are unevenly distributed across geography and practice area. According to the ABA, “[l]awyer discipline rates vary significantly from state to state” (2022, p. 85). In Alabama and Iowa, 1 out of every 100 lawyers received public discipline in 2019; in the same year, the rate of discipline in Alaska and Rhode Island was 10 times lower. The variability in these rates undermines the reliability of the measure. Additionally, most instances of discipline involve attorneys who are engaged in solo or small private practice (Levin et al., 2013; State Bar of California, 2019). This suggests that lawyers who serve individual clients (e.g., private practice lawyers) are more likely to have complaints than lawyers who represent corporate or public entities and lawyers who do not have clients at all.

Finally, discipline rates vary by gender, age, and race, perhaps due in some part to racial and gender representation among practice areas (National Association for Law Placement [NALP], 2022). Hatamyar and Simmons (2004) find that (1) men are more likely to be disciplined than women, and (2) disciplined attorneys tend to be older than non-disciplined attorneys on average. And the State Bar of California (2019) reports that compared to White male attorneys, Black male attorneys in California were more likely to be placed on probation (1.4% vs. 0.9%) or disbarred (1.6% vs. 1.0%), when controlling for the number of complaints.

In summation, the limitations of using lawyer discipline rates as a measure of competence may qualify the claim that there is a meaningful relationship between bar exam scores or cut scores and lawyering competence. Discipline is a useful but inherently limited proxy for competence to practice law—and achieving a more nuanced understanding of its

relationship to bar exam passage or bar exam scores might require further study.

Study Design

Theoretical Framework

For the bar exam to be a sound instrument used to measure lawyering competence and therefore minimum competency, the interpretations of its results should be valid. According to the Standards for Educational and Psychological Testing (published by the American Educational Research Association, the American Psychological Association, and the National Council on Measurement in Education), validity in standardized testing is “the degree to which evidence and theory support the interpretations of test scores for proposed uses” of a given test (American Educational Research Association et al., 2014, p. 11). Although it is not uncommon to hear standardized tests or test scores spoken of as “valid” or “invalid,” this is not consistent with how the term is used in contemporary psychometrics: “a measure itself is neither valid nor invalid; rather, the issue of validity concerns the interpretations and uses of a measure’s scores” (Furr, 2018, p. 220). There are many aspects of validity in standardized testing, but the most fundamental is construct validity: the extent to which a measure’s scores may be interpreted as indicative of the presence or intensity of a psychological construct (Messick, 1989).

We therefore focus our attention on the question of construct validity with respect to minimum competence to practice law: whether a bar exam score may defensibly be interpreted as predicting that a candidate will demonstrate the competencies needed to be an effective lawyer. We test this relationship by focusing on the concept of lawyering effectiveness, as developed by Shultz and Zedeck (2011). As discussed above, Shultz and Zedeck not only identified 26 competencies important to lawyering effectiveness but also developed and validated a means of measuring each using a BARS methodology. The authors granted permission for these scales to be used in our study to measure lawyering effectiveness (see Data Collection Process below).

We therefore offer an empirical assessment of whether, and to what extent, first-time bar exam scores and passage are predictive of lawyering effectiveness. In other words, we are interested in investigating whether candidates with higher bar exam scores are likelier to exhibit the traits associated with effective lawyering. We rely on a measure of lawyering effectiveness, the outputs of which have a strong presumption of being validly interpreted as indicating aspects of lawyering effectiveness: Schultz and Zedeck’s scales. If the bar exam is a meaningful measure of minimum competence, then we would expect to see a positive relationship between bar exam scores and rating of lawyering effectiveness. Given the racial disparities in bar passage rates discussed above, any

relationships between exam scores and lawyering effectiveness should be compelling and consistent.

Data

Our sample comprises 524 lawyers who were admitted to the Nevada Bar between 2014 and June 2020 and who had not previously been admitted to a state bar in another jurisdiction.³ It was drawn from the population of 1414 such new lawyers in Nevada. Participation was voluntary but incentivized via a gift card offer and the awarding of continuing education credits. There are only modest differences between participants and nonparticipants regarding their mean MBE and essay scores (and their standard deviations) and the proportion that took the bar exam more than once (see Table 1). The average MBE score differs by less than one point, the average essay score differs by about one point, and the proportion of those taking the bar exam more than once differs by less than four percentage points. None of these differences are statistically significant.

Data Collection Process

This sample was achieved via a three-stage process. First, we obtained the names and contact information for those 1414 “new lawyers” who were admitted to the Nevada Bar and sat for the Nevada Bar Exam from 2014 through the February 2020 administration. For each of these new lawyers, we received their first-time MBE and MPT scores, as well the scores they received on the Nevada subject essays and the ethics essay components. We also were able to ascertain whether the attorney took the Nevada Bar Exam more than once.

Next, beginning in September 2021, we contacted these 1414 new lawyers via email, soliciting their participation in the study. This involved directly emailing the population of new Nevada lawyers using the survey program Qualtrics[®], asking for their consent to participate and to provide information regarding their demographic background and current employment and specializations. In addition, we asked them to provide contact information for two supervisors, two peers, and one judge who would be qualified to assess their lawyering effectiveness. We also asked them to complete an

optional five-item self-assessment of lawyering effectiveness. These five items were a subset of the 26 identified by Shultz and Zedeck.

To incentivize study participation, the State Bar of Nevada offered participating new lawyers a \$10 gift card and 3 credits toward their annual Continuing Legal Education (CLE) requirements. We received usable responses from 524 of the lawyers (hereinafter referred to as “participants”)—a response rate of 37%.

Last, in October and November 2021, we contacted the participants’ supervisors, peers, and judges, as provided in their surveys. These external raters were asked to verify their working relationship with the participant and to evaluate the participant’s lawyering effectiveness on the 26 factors published by Shultz and Zedeck.⁴ Virtually all participants (453 of 524; 86.4%) received at least one evaluation from a supervisor, peer, or judge, with an average of 1.64 external reviews per participant.⁵ Table 2 indicates the number of attorneys for whom we received evaluations by evaluator type.

Variables

Outcomes. Our outcome of interest is lawyering effectiveness, as measured by self-assessment and by external raters. To reduce survey burden, we asked participants to self-assess their abilities across only 5 (of the 26 identified by Shultz and Zedeck) items that we considered to be most broadly relevant and applicable areas across a multitude of practice areas and settings (e.g., solo practitioner, large firm). The five items included in the self-assessment battery were:

- Ability to use analytical skills, logic, and reasoning to approach problems and to formulate conclusions and advice.
- Understanding of legal concepts and utilizing sources and strategies to identify issues and derive solutions.
- Ability to identify relevant facts and issues in a case.
- Ability to generate well-organized methods and work products.

Table 2. Number of Attorneys for Whom Evaluations Were Received, by Rater Type.

Evaluation Type (#)	Participants (%) (N = 524)
Self-assessment	460 (88)
External (≥1)	470 (90)
Supervisor (≥1)	227 (43)
Supervisors (2)	150 (29)
Peer (≥1)	221 (42)
Peers (2)	111 (21)
Judge (1)	149 (28)

Table 1. Comparison of Study Participants and Non-participants.

Total Eligible Lawyers Solicited for Participation (N = 1414)		
Participants (n = 524)		Non-Participants (n = 901)
148.9 (10.8)	Avg. MBE score (SD)	148.0 (10.7)
711.5 (38.2)	Avg. Essay score (SD)	714.9 (41.6)
15.7%	Took bar exam more than once	19.2%

- Ability to write clearly, efficiently, and persuasively.

Each question included a 1–5-point scale, with half-point increments, and an accompanying list of behavioral anchors (examples of behaviors and activity that demonstrate each level along the scale). These behavioral anchors were adapted from those Shultz and Zedeck created, utilized, and tested.

Responses to the self-assessment were not required, yet 469 of the 524 participants (90%) answered at least one of the self-evaluation items. Of these 469 participants, 456 (97 %) answered all five questions. And of the 13 participants that did not answer all of the questions, one answered two questions, one answered three questions, and 11 answered four questions. For each participant, we calculated their average self-assessment rating. Of the 469 participants, 99% had average self-assessment ratings between 3.0 and 5.0 points. Across the 469 participants, the mean self-assessment rating was 4.20 (SD = 0.44), with a minimum of 2.5 and a maximum of 5.0 (see Table A.3).

Each of the five items has a positive and statistically significant relationship with the others, meaning new lawyers who rated themselves highly on one factor tended to also rate themselves highly on the others (see Table A.4). Given these intercorrelations, we performed a principal components analysis to test whether the five items might be best captured by one latent variable. Our results suggest that combining the scores across the five self-assessment items would best represent the variance in self-ratings (see Table A.5).

Each supervisor, peer, and judge identified by the study participants was asked to provide demographic and background information as well as an evaluation using a uniform 26-item battery. These 26 items corresponded to those skills identified and validated by Shultz and Zedeck, which they categorized by theme (see Table A.2). Each item included a set of behavioral anchors (also developed by Shultz and Zedeck) that describe how a lawyer demonstrates each level of the rating scale.

As shown in Table 3, there is clear variation in score averages across the eight skills themes as well as across the different types of raters.⁶ Supervisors rated the participants lower than peers and judges across all eight categories. Moreover, it is noteworthy that the average self-assessment was 4.20—lower than any of the external reviewers. This suggests that self-evaluation responses are not inflated.

Each of the eight skill categories has a positive and statistically significant relationship with the others, meaning a reviewer who rated a participant highly in one area tended to rate them highly on the others as well (see Table A.6). Relative to the self-evaluation scores, the intercorrelations for external reviewers are considerably stronger. This trend suggests that each of the skill categories captures the same underlying dimension—lawyering effectiveness. As with the self-assessment, we performed principal factor analysis to determine whether the 26 items might load onto fewer latent factors. We found that all 26 items loaded onto one factor for each reviewer type (see Table A.7).

We determined that we would not combine scores across rater types after examining the correlations between the self-evaluation and the overall supervisor, peer, and judge evaluations, both overall and by each thematic category. Each of the evaluation types has only a modest positive but statistically significant relationship with the others (see Table A.8). These relatively low intercorrelations among raters could suggest real divergence in perceptions as well as differences in response styles among the three types of raters. With respect to response styles, each of the raters may differ in their interpretation of the wording in individual questions, or the extent to which they strictly apply the behavioral anchors when choosing a rating. Additionally, supervisors, peers, and judges likely vary in the depth of their experiences with participants across the eight thematic categories.

In the end, we have one latent variable measuring lawyering effectiveness for each reviewer type (i.e.,

Table 3. Average Evaluations From External Evaluators and Self-Evaluations.

Category	Supervisor(s)	Peer(s)	Judge	Self
Intellectual & cognitive factors	4.20	4.39	4.32	4.29
Research & info gathering	4.22	4.35	4.29	4.22
Communications	4.14	4.31	4.36	4.29
Planning & organizing	4.14	4.32	4.36	4.11
Conflict resolution	4.20	4.37	4.34	— ^a
Client & business relationships	4.16	4.39	4.41	— ^a
Working with others	4.33	4.47	4.63	— ^a
Character	4.35	4.47	4.49	— ^a
Overall	4.22	4.38	4.38	4.20 ^b

^aParticipants were not given self-assessment items related to this skill.

^bAverage for self-evaluation includes the “Knowledge and the Law” factor, which was not included in external evaluations.

self-assessment, peer reviewer, supervisor reviewer, judge reviewer).

Explanatory Variables. Our explanatory variables represent different aspects of new lawyers' first-time performance on the Nevada Bar Exam. By focusing on first-time performance, we are able to explore a wider range of scores, including those below the cut point.

During the study period, the Nevada Board of Bar Examiners made use of the MBE and MPT as two components of its bar exam.⁷ In place of the MEE, Nevada administered its own jurisdiction-specific seven-question essay examination. These essay scores were scaled to the MBE and then combined to determine an examinee's total bar score and, ultimately, whether they passed.⁸ Unlike the MEE, one of the essay topics focuses on legal ethics, in addition to a similar list of possible subjects tested on the MEE.

We utilize new lawyers' overall first-time performance on the bar exam, as well as their performance on each of the individual components of the exam to explain ratings of lawyer effectiveness as measured by self-evaluations and evaluations from peers, supervisors, and judges. We measure overall performance as both a binary result (pass or fail) and as a combined score. To derive the combined score, we use the same method that the NCBE uses to calculate a score for the UBE (the sum of 50% scaled MBE score, 30% scaled subject essay score, and 20% scaled MPT score).

We also use the scores on each of these individual components to model ratings of lawyer effectiveness. In addition, we also use the score on the specific ethics essay question as an explanatory variable because this information was provided for each observation in our sample. Each explanatory variable is scaled such that it has a minimum of zero and a maximum of one.

Control Variables. We consider several possible control variables, including the new lawyers' racial/ethnic and gender identification, law school attended, practice area and setting, years since taking the bar exam, age, and survey wave. The latter refers to whether or not the individual was part of the soft launch of the survey, which we used to test the logistics of the survey instrument and our distribution strategy, or the full launch. Ultimately, the decision to include a particular control variable was determined by whether its inclusion improved model fit, namely using the Bayesian Information Criterion, or BIC.

In addition, we consider the interplay between the gender and race/ethnicity (among other characteristics) of the evaluators and the ratings they provide and for whom. Although we do find that average ratings vary to some extent based on the race/ethnicity of the evaluator (but not their gender), we do not include evaluator characteristics among our control variables. We do this because evaluators provided their ratings after the attorneys took and passed

the bar exam, which means that their rating did not affect the attorney's bar exam performance. This independence means that the characteristics of the evaluators cannot bias the estimates of the relationships we explore between an attorney's bar exam performance and their rating of lawyer effectiveness. To confound or bias our estimates, a variable would need to be related to both the outcome (ratings of lawyering effectiveness) and our explanatory variable (bar exam performance). Thus, although interesting, the interplay between the attributes of the evaluators and the attorneys they rate are not useful control variables in this instance.

Analytical Approach

Before creating detailed models analyzing our predictors and outcomes of interest, we first undertook a descriptive exploration of each relationship using locally estimated scatterplot smoothing (LOESS) plots and bivariate ordinary least squares (OLS) models. The LOESS figures allow us to see the relationships between various bar exam scores and lawyering effectiveness ratings without any assumptions regarding linear effects. In each case, we find that the relationships are effectively linear. [Figure 1](#) illustrates the LOESS curve for the relationship between first-time bar exam score and ratings of lawyering effectiveness, by evaluator type. The relative linearity of this relationship is representative of the relationships we see between ratings of lawyering effectiveness and the other measures of bar exam performance (see [Figures A.1–A.4](#)). We use the OLS models to measure how much variation each bar performance indicator explains within each rating outcome.

We then construct Tobit regression models to quantify the relationships between each bar exam performance indicator of interest and each type of lawyering effectiveness rating. The use of Tobit models improves our ability to navigate constraints on the range of lawyering effectiveness scores reported. For example, although peer raters could provide ratings ranging from 1 to 5, no peer rater provided a rating lower than 2.5. Since we collected ratings data from four groups, this approach results in four distinct models per bar performance predictor (self-evaluation, peer evaluation, supervisor evaluation, and judge evaluation).

The level of censoring at the left tails of the modeling distributions varies somewhat across models given the tendencies of certain rater groups to award higher or lower scores. Therefore, all self-evaluation and peer evaluation models are censored below a score of 2.5. Meanwhile, all supervisor evaluation models are censored below a score of 1.5, and all judge evaluations are censored below a score of 1. (Note that the results from our Tobit models closely resemble those attained using OLS regression).

Furthermore, to account for heteroskedasticity in the residuals, we report robust standard errors. This allows us to apply a

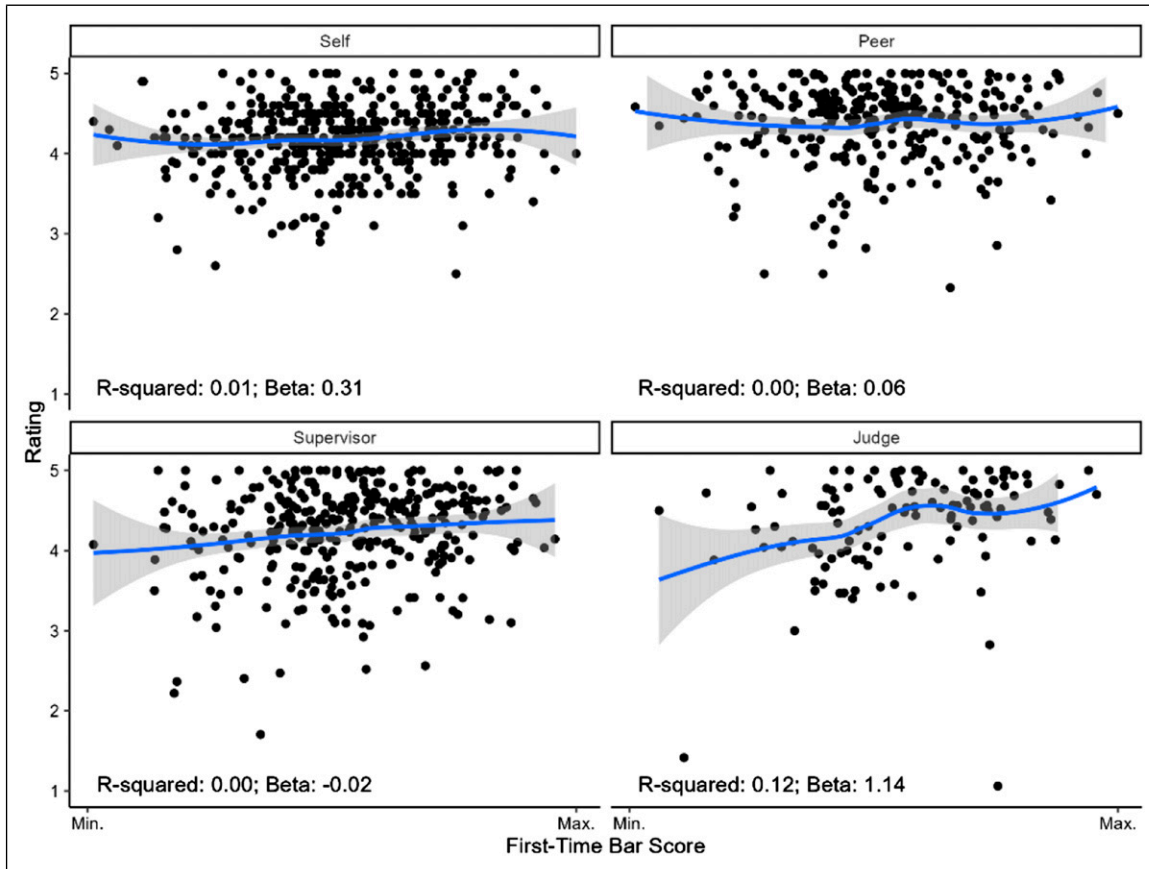


Figure 1. LOESS curves: Lawyering effectiveness rating by bar score.

correction to the level of precision (our standard errors) in our estimates.

Findings

In this section, we explore the effects of first-time bar exam performance (both overall and on the individual components of the exam) on ratings of lawyer effectiveness. When we discuss the relationships between bar exam performance and lawyer effectiveness ratings, we do so in terms of percentage changes. Since our outcomes are on a scale from 1 to 5, we divide the estimated *point* change in lawyer effectiveness by 5 to calculate a *percent* change. The results discussed below control for attorney age, gender, and race/ethnicity, as well as survey wave, and time elapsed since the attorney sat for the bar exam.

Overall Bar Exam Performance

As noted above, we measure overall first-time bar exam performance as both a binary result (pass/fail) and as a combined score. Simple bivariate OLS models reveal that neither first-time bar passage nor first-time bar score alone explain more than 12% of the variation in any of the lawyer effectiveness ratings (see [Tables A.9 and A.10](#)). For each of these explanatory models, we use four

Tobit regression models, one for each type of evaluator (i.e., self, supervisor, peer, and judge), to estimate the magnitude of the relationships between overall bar exam performance and lawyer effectiveness. This allows us to control for the censoring of the data.

Holding all else equal, we find that those who passed on their first attempt have approximately 1% higher self-ratings, 1% higher peer ratings, 5% higher supervisor ratings, and 6% higher judge ratings. Only the effect of first-time bar passage on supervisor ratings is statistically significant ($p < .05$).

When we consider first-time bar score, we find that a one standard deviation increase in bar exam score is associated with an approximate 1% increase in self-rating, no change in peer rating, a 2% increase in supervisor rating, and a 4% increase in judge rating. Each of these effects is statistically significant, except for that between peer rating and bar exam score.

Individual Exam Components

MBE Scores & Lawyering Effectiveness. Next, we examine the extent to which increases in Nevada lawyers' first-time MBE scores influence the lawyering effectiveness ratings they receive from peers, supervisors, judges, and themselves. The bivariate OLS models reveal that first-time MBE score

explains no more than 10% of the variation in any rating of lawyer effectiveness (see [Table A.11](#)).

The Tobit regression models suggest that first-time MBE score has a weakly positive association with lawyering skills ratings. We find that, holding all else constant, a one standard deviation increase in MBE score is associated with an approximate increase in ratings from self-evaluations of 2%, peer evaluations of 0%, supervisor evaluations of 2%, and judge evaluations of 3%. The effects of first-time MBE score on self-evaluation and supervisor evaluation ratings are statistically significant ($p < .05$).

MPT Scores & Lawyering Effectiveness. We also investigate the extent to which higher first-time MPT scores influence the lawyering effectiveness ratings that peers, supervisors, judges assign new Nevada lawyers—as well as the ratings Nevada lawyers assign to themselves. Our simple bivariate OLS models reveal that first-time MPT score explains no more than 3% of the variation in lawyering effectiveness rating (see [Table A.12](#)).

Our estimates from the Tobit models suggest that first-time MPT score does not have a practically significant effect on predicted lawyering skills evaluation score from any of the four rater groups: self, peer, supervisor, or judge. Holding all else constant, a one-standard deviation increase in MPT score is associated with an approximate increase in ratings from self-evaluations, peer evaluations, and supervisor evaluations of 0%, and an increase in judge evaluations of 3%. None of the effects are statistically significant ($p < .05$).

Nevada Subject Essay Scores & Lawyering Effectiveness. Next, we examine the extent to which increases in Nevada lawyers' first-time subject essay scores influence the lawyering effectiveness ratings they receive from peers, supervisors, judges, and themselves. The bivariate OLS models reveal that subject essay score explains no more than 2% of the variation in any rating of lawyer effectiveness (see [Table A.13](#)).

The Tobit regression models suggest that subject essay score has a weakly positive association with lawyering skills ratings. We find that, holding all else constant, a one standard deviation increase in subject essay score is associated with an increase in ratings from self-evaluations of 1%, peer evaluations of 0%, supervisor evaluations of 0%, and judge evaluations of 3%. Only the relationship between first-time subject essay score and judge evaluation ratings is statistically significant ($p < .05$).

Ethics Essay & Lawyering Effectiveness. Finally, we investigate the relationships between first-time legal ethics essay scores and lawyering effectiveness ratings from each rater type. The legal ethics essay constitutes just one section of the overall Nevada subject essays component of the bar exam analyzed above. The bivariate OLS models reveal that ethics essay score explains no more than 7% of the variation in any rating of lawyer effectiveness (see [Table A.14](#)).

The Tobit regression models suggest that ethics essay score has a weakly positive association with lawyering skills ratings. We find that, holding all else constant, a one standard deviation increase in ethics essay score is associated with an increase in ratings from self-evaluations of 0%, a decrease in peer evaluations of 1%, an increase in supervisor evaluations of 1%, and an increase in judge evaluations of 2%. None of the effects are statistically significant ($p < .05$).

[Table 4](#) summarizes the change in each type of lawyering effectiveness rating associated with a one-standard deviation improvement in each bar performance measure analyzed.

Overall, improvements within each facet of Nevada bar exam performance (passage, overall score, or section scores) evidently correspond to only negligible gains—or occasional decreases—in lawyering effectiveness ratings. The effects, although quite weak, are generally strongest when overall exam score is analyzed as the predictor or when judge rating is analyzed as the outcome. Otherwise, however, the models and descriptive analyses suggest weak or absent relationships between bar performance and lawyering effectiveness ratings.

Discussion

Given the bar exam's stated purpose of identifying candidates with the "minimum competence" to practice law, and its assumed pass-rate disparities for law school graduates of color, the question of whether or not bar exam scores may be validly interpreted as evidence of "minimum competence" warrants scrutiny. Moreover, given the paucity of validity evidence, the burden of proof should lie with the bar exam. Bar exam scores should have meaningful and positive relationships with the skills required for effective lawyering—in this case, the 26 lawyering skills developed by Shultz and Zedeck.

As we show above, assessments of lawyering effectiveness using Shultz and Zedeck's scales should be treated as presumptively valid. Using principal factor analysis and confirmatory factor analysis testing, we find strong evidence that for each type of rater, taken collectively, the 26 lawyering skills measure the same latent variable—lawyering effectiveness. It appears unlikely that self-assessments are inflated, since average self-assessments are lower than any of those provided by external reviewers. Any differences observed in ratings between supervisors, peers, and judges are likely a reflection of their differing relationships to, and interactions with, the study participants.

Our results suggest that the bar exam does not meaningfully predict the demonstration of the competencies to be an effective lawyer. We find only modest, positive relationships between first-time bar exam scores and all four assessments of lawyering effectiveness (i.e., self, supervisor, peer, and judge). These effect sizes are not practically significant. Holding constant the attorney's age, gender, race/ethnicity, and time elapsed since sitting for the bar exam, as

Table 4. Effect of a One-Standard Deviation Increase in Bar Score on Lawyering Effectiveness, by Bar Score Metric and Rater Type.

Score (+ 1σ)	Expected Percent Change (%) in Lawyering Effectiveness Score by Rater			
	Self	Peer	Supervisor	Judge
Overall	1.4	0.5	2.1	4.2
MBE	1.7	0.2	1.6	3.0
MPT	0.2	0.0	0.5	2.7
Subject Essays	1.0	-0.5	0.5	3.2
Ethics Essay	-0.2	-0.7	0.7	1.6
Bar Passage*	1.0	0.9	5.2	6.5

Note: These effects are estimated using Tobit regression. Bold effects are statistically significant. *Since bar passage is not a continuous predictor variable, reported effects are associated with the binary outcome of passing the exam rather than with a one-standard-deviation change.

well as the survey wave in which they participated, among new lawyer in Nevada, the largest effect we see for a one-standard-deviation increase in score on the bar exam is a 4% increase in rating of lawyering effectiveness by judges. The results are mixed for the MPT, Nevada subject essay, and Nevada ethics essay scores—depending on the rater, the effects may be either positive or negative. But in nearly all cases, the effects do not have practical significance. The only exception is the relationship between subject essay scores and judge’s ratings: lawyers that scored higher on the subject essays were rated higher. Although these effect sizes are the largest that we find, their standard errors indicate that there is substantial uncertainty around these estimates.

Based on these findings, we conclude that the bar exam—as it was administered in Nevada—does not result in scores that serve as bases for valid inferences about predicted lawyering effectiveness. Although some results achieve statistical significance, this threshold should not be used as the sole determining factor as to whether the bar exam serves its purpose. Several factors are involved when estimating *p*-values (e.g., sample size, statistical power), which can render statistical significance for even exceptionally small effects. Although *p*-values are a useful tool, they should never be viewed without context—that is, interpretation of the coefficient and the practical significance of the result. Therefore, the bar exam’s use to determine who possesses the minimum competence to practice law and gain admittance to the bar should be questioned and subjected to further rigorous study.

The UBE is scheduled for a major revision beginning in 2026, and therefore it may be tempting to devalue or dismiss these findings. However, the NextGen Bar Exam will retain variations of the MBE and MPT and thus to some extent our findings will apply to the next iteration of the bar exam. But the larger lesson is that the validity of interpretations of bar exam scores, in whatever format the exam takes, has been insufficiently vetted. While the NextGen Bar is being designed, the NCBE should be actively engaged in rigorously assessing both the constructs that the test is intended to

measure and the content of that test. Furthermore, the results of such assessments should be used to provide public, clear guidelines informing decisions surrounding cut scores, or what constitutes “minimum competence.”

Further study should additionally focus on identifying other factors that account for the acquisition of skills and noncognitive factors that conduce to lawyering effectiveness. Elapsed time since bar passage is one such possibility—we find that it is positively associated with lawyering effectiveness, suggesting that as lawyers gain more experience, their effectiveness increases. However, our analyses do not permit us to make any definitive claims because this variable is included only as a control variable and as such its purpose it to estimate more accurately the magnitude of the relationships between bar exam scores and ratings of lawyering effectiveness. Any inferences based on time since bar passage should be avoided because, as a control variable, we do not include other potential confounders that might bias the size of its relationship with lawyering effectiveness. Nonetheless, this finding might suggest that further research, including whether the adoption of a medical school model with supervised practice, is warranted. Perhaps this experience might ultimately be the best predictor of effective lawyering—evidence suggests that residency is a particularly effective training technique for new medical doctors. This is consistent with both the literature and the totality of our findings: noncognitive factors—not easily assessed by a multiple-choice instrument like the MBE, or even constructed-response instruments like the MPT—play a crucial role as determinants of lawyering effectiveness.

A group of practitioners commissioned by the Nevada Supreme Court recently submitted a series of recommendations to restructure the bar exam in Nevada, including the introduction of a residency-type requirement. In brief, the initial recommendations of the commission are to require bar applicants to:

- Following a student’s completion of their first year of curriculum, pass a knowledge-based, multiple-choice

exam based on the doctrinal coursework taken during this time.

- Following graduation, successfully complete a series of practice-based essay questions (the Nevada Performance Test), which are developed by the State Bar and which are scored based on a Nevada-created rubric.
- During their third year or following graduation, successfully complete a period of supervised practice.

Our results support these recommendations in several ways. First, we find that MBE scores are only minimally predictive of lawyering effectiveness. The topics tested on the MBE are typically those covered in law students' first year of studies, which might partly explain its weak relationship with lawyering effectiveness. By testing this acquisition of this knowledge closer to the time at which students learn this material, this component of the bar exam might be a better predictor. In addition, the commission has recommended that those who fail the knowledge test not be allowed to continue their studies until they pass. Such timing would also allow students an earlier chance to gauge their potential for bar admission. This would mean that students could change their career or academic path after only one year. Currently, waiting to take the bar exam until after graduation means that those who fail will carry three (or more) years of law school debt without the additional income concomitant with bar licensure.

Second, we find that the MPT is negligibly related to lawyering effectiveness. It is possible that the prompts created by the National Conference of Bar Examiners do not capture the practice of law that is specific to Nevada or that the rubric used to grade the responses are not aligned with best practices in Nevada. The adoption of the Nevada Performance Test may allow for better tailored essay prompts and answer keys that more closely align with practice in Nevada.

Last, as we note above, the control variable years since passage might suggest that some form of supervised practice could yield more effective lawyers. If years since passage captures the experience of practicing, it is reasonable to assume that with greater practice comes greater effectiveness. It could therefore be a significant boost to lawyering effectiveness to require this type of supervised practice, which is similar to the training of medical doctors.

Additional study when the recommendations have been implemented will be necessary to determine whether the new format better measures and therefore better predicts lawyering effectiveness compared to the administration of the bar exam studied here.

Limitations

As with any empirical study, there are limitations borne out of the data we have available to us for analysis. All analyses of the effectiveness of credentialing or admissions in education

suffer from the same significant inferential issue—censored data. Specifically, while we can observe variation in lawyering effectiveness across the full range of lawyers who eventually passed the bar exam, we cannot observe the potential lawyering effectiveness of those who never successfully became new lawyers.

To some extent, we are able to account for this limitation by using first-time bar exam scores. Of the 524 participants, 84 (16%) failed the exam on their first attempt; thus, we are able to examine how the bar exam performs in predicting lawyering effectiveness for those individuals scoring below the cut score. Overall, approximately half of those who failed the bar exam on their first attempt received lawyering effectiveness ratings at or above the average for each evaluation type. Thus, it seems that the bar exam is no more predictive of lawyering effectiveness for those that fail on their first attempt than it is for those that pass.

Nevertheless, it is important to recognize that the participants in our sample did eventually pass the bar and enter the legal profession in Nevada. There are some applicants to the bar who never achieve a passing score. We are unable to take these individuals into consideration in these analyses as we can never know how effective they would have been as lawyers. As such, we do not know the extent to which these individuals differ from those who failed on their first attempt but ultimately passed. These differences could lead to biased estimates, but we cannot know or approximate the direction or the magnitude of the potential bias. It is possible that bar exam scores are more impactful among those that never pass the exam. But it is also possible that they are less so.

Additionally, our ratings of lawyering effectiveness are effectively limited to a range of three points to five points. This data censoring means that there is little room for sizable increases or decreases in ratings as a function of bar exam score. Nonetheless, even when considered relative to the narrower range of rating values, the magnitude of the predictive effect of bar exam scores is small. An argument could be made that the bar exam is responsible for the floor on rating values—that is, those that pass the bar exam are effective lawyers, receiving at least a three on the ratings scales. But, as noted two paragraphs above, this argument is contrary to the data we have for those who failed on their first attempt. Of these individuals, approximately half score at or above the average rating, suggesting that those that fail the bar exam are no less effective.

We also only observe bar performance and lawyering effectiveness in this current study, and do not know how individual metrics that predate bar performance, such as performance in law school, law school admissions metrics, or other characteristics might relate to the outcome of lawyering effectiveness. Future research on the topic should explore what factors *do* predict lawyering effectiveness.

Finally, our sample comprises only those lawyers admitted to the Nevada Bar. It is unclear the extent to which our sample

is representative of the population of new lawyers during the study period and therefore whether they generalize to the national population.

Conclusion

Passage of a bar exam is required in nearly all United States jurisdictions in order to practice law. Theoretically, this is because bar results indicate how well a burgeoning attorney will fare in their career, with higher scores indicating a promising future, and those below a particular cut point signaling lack of skills required for minimum competency. As such, the bar exam should, presumably, significantly predict ratings of lawyering effectiveness. Thus, our analysis builds on Schultz and Zedeck's earlier work to test whether the relevant bar scores are related to their 26 skills of lawyering effectiveness. We find that while MBE scores are statistically significantly related to some evaluations of lawyering effectiveness, the relationships are small and offer little practical significance. Additionally, the majority of variance remains unexplained, even when including relevant control variables. These findings suggest that the bar exam, as it was administered in Nevada to new lawyers during the study period, may not be the indicator of future career effectiveness for new attorneys that it should be, particularly given equity concerns. More research is needed, but this study finds that while the bar is serving as a significant barrier to the practice of law, there is little indication that it is a robust indicator of what it takes to be a "good" lawyer.

Declaration of Conflicting Interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) received no financial support for the research, authorship, and/or publication of this article.

ORCID iD

Stephen N. Goggin  <https://orcid.org/0000-0001-8653-617X>

Supplemental Material

Supplemental material for this article is available online.

Notes

1. Wisconsin does not require a bar exam for graduates of either law school in the state: University of Wisconsin Law School and Marquette University Law School. Law school graduates from other states wanting to practice law in Wisconsin must pass the Wisconsin Bar Exam, have passed one in a state with which it has an agreement of reciprocity, or be granted admission on proof of practice. In New Hampshire, graduates of the University of New Hampshire School of Law's Daniel Webster Scholar Honors

Program are eligible for admission to the New Hampshire Bar without further examination.

2. In fact, [Schemenauer \(2007\)](#) surveyed the cause of complaints in all but 10 jurisdictions in 2007, finding that complaints related to a failure to communicate were among the most common in the majority of jurisdictions. Yet, client communication is not a skill tested on the bar exam.
3. Those admitted in June 2020 would have taken the bar exam in February 2020, prior to the widespread disruptions caused by the COVID-19 outbreak.
4. "Working relationship" was defined broadly as time spent time in a professional capacity working with, observing, or otherwise interacting with the study participant in such a way that the evaluator had first-hand experience/knowledge of the participants professional abilities and skills.
5. Some evaluators provided assessments for multiple study participants. This was most notable among judge evaluators.
6. Although study participants were not given the full battery of 26 skills, they were given a subset of 5 of these items, each of which falls under one of the 8 categories. Thus, we are able to compare participants' self-evaluations to the average scores they received from supervisors, peers, and judges for these 5 common items.
7. The Nevada Supreme Court has suspended use of the MBE for several consecutive administrations as a result of the COVID-19 pandemic ([State Bar of Nevada, 2022](#)). The Nevada Supreme Court has now ordered that the MBE be used as a knowledge exam and not used to scale the essay questions and the Nevada Performance Test (NPT) essay scores. Beginning in 2020, the Nevada Supreme court replaced the MPT component of the bar exam with the NPT, which are a set of practice-based essay prompts created by and scored based on a Nevada-created rubric.
8. In Nevada, and in virtually every jurisdiction, scaling is used to minimize the risk that examinees are unfairly penalized by variation in exam difficulty or grader leniency. The procedure involves standardizing, separately, the distribution of the MBE and essay scores around its mean with a standard deviation of one. Each examinee's score in standard deviation units is then multiplied by the standard deviation of the scaled MBE scores in the jurisdiction ([Case, 2006](#)).

References

- American Bar Association. (1992). Legal education and professional development—An educational continuum, Report on the Task Force on Law Schools and the Profession: Narrowing the Gap. [MacCrate Report].
- American Bar Association. (2022). *ABA Profile of the legal profession 2022*. <https://www.americanbar.org/content/dam/aba/administrative/news/2022/07/profile-report-2022.pdf>
- American Educational Research Association, American Psychological Association, National Council on Measurement in Education. (2014). *Standards for educational and psychological testing*.

- Anderson, R. IV, & Muller, D. T. (2019). The high cost of lowering the bar. *The Georgetown Journal of Legal Ethics*, 32, 307.
- Benthall-Nietzel, D. (1974). An empirical investigation of the relationship between lawyering skills and legal education. *Kentucky Law Journal*, 63(2), 373.
- Case, S. (2006). Frequently asked questions about scaling written test scores to the MBE. *The Bar Examiner*, 75(4), 42–44. <https://thebarexaminer.ncbex.org/wp-content/uploads/PDFs/750406-Testing.pdf>
- Foster, S. (2021). Does the Multistate Bar Exam validly measure attorney competence? *Ohio State Law Journal Online*, 82, 31.
- Furr, R. M. (2018). *Psychometrics: An introduction*. Sage Publications.
- Hatamyar, P. W., & Simmons, K. M. (2004). Are women more ethical lawyers— An empirical study. *Florida State University Law Review*, 31(4), 785.
- Howarth, J. W. (2017). The case for a uniform cut score. *Journal of the Legal Profession*, 42(1), 69–84.
- Kinsler, J. S. (2017). Is bar exam failure a harbinger of professional discipline. *St. John's Law Review*, 91(4), 883.
- Levin, L. C., Zozula, C., & Siegelman, P. (2013). A study of the relationship between bar admissions data and subsequent lawyer discipline. Available at SSRN 2258164.
- Merritt, D. J., Hargens, L. L., & Reskin, B. F. (2001). Raising the bar: Social science critique of recent increases to passing scores on the bar exam. *University of Cincinnati Law Review*, 69(3), 929–968. <https://doi.org/10.2139/ssrn.249794>
- Messick, S. (1989). Validity. In R. L. Linn (Ed.), *Educational measurement* (pp. 13–103). American Council on Education.
- National Association for Law Placement. (2022). *Jobs & JDs: Employment and salaries of new law graduates – Class of 2021*.
- National Conference of Bar Examiners. (2021). *Overview of recommendations for the next generation of the bar exam*. <https://nextgenbarexam.ncbex.org/overview-of-recommendations/>
- Patton, W. W. (2020). Admitting law graduates by bar examination versus by diploma privilege: A comparison of consumer protection. *Journal of the Legal Profession*, 45, 243. <https://doi.org/10.2139/ssrn.3686252>
- Plomin, R. (2023). Celebrating a century of research in behavioral genetics. *Behavior Genetics*, 53(2), 75–84. <https://doi.org/10.1007/s10519-023-10132-3>
- Plomin, R., Shakeshaft, N. G., McMillan, A., & Trzaskowski, M. (2014). Nature, nurture, and expertise. *Intelligence*, 45(100), 46–59. <https://doi.org/10.1016/j.intell.2013.06.008>
- Rosin, G. S. (2008). Unpacking the bar: Of cut score and competence. *Journal of the Legal Profession*, 32, 67–98.
- Schemenauer, S. E. (2007). What we've got here.. is failure.. to communicate: A statistical analysis of the nation's most common ethical complaint. *Hamline Law Review*, 30(3), 629–682.
- Schwartz, R. A. (1973). The relative importance of skills used by attorneys. *Golden Gate University Law Review*, 3(2), 321.
- Shultz, M. M., & Zedeck, S. (2011). Predicting lawyer effectiveness: Broadening the basis for law school admission decisions. *Law & Social Inquiry*, 36(3), 620–661. <https://doi.org/10.1111/j.1747-4469.2011.01245.x>
- State Bar of California. (2019). *Report on disparities in the discipline system*. <https://board.calbar.ca.gov/docs/agendaItem/Public/agendaitem1000025090.pdf>
- State Bar of Nevada. (2022). *July 2022 bar exam modified to exclude MBE*. <https://nvbar.org/july-2022-bar-exam-modified-to-exclude-mbe/>
- Wadsworth, S. J., Corley, R. P., Hewitt, J. K., Plomin, R., & DeFries, J. C. (2002). Parent-offspring resemblance for reading performance at 7, 12 and 16 years of age in the Colorado Adoption Project. *Journal of Child Psychology and Psychiatry*, 43(6), 769–774. <https://doi.org/10.1111/1469-7610.00085>
- Winick, M., Quintanilla, V. D., Erman, S., Chong-Nakatsuchi, C., & Frisby, M. (2020a). Examining the California cut score: An empirical analysis of minimum competency, public protection, disparate impact, and national standards. AccessLex Institute Research Paper.
- Winick, M., Quintanilla, V. D., Erman, S., Chong-Nakatsuchi, C., & Frisby, M. (2020b). A five-year retroactive analysis of cut score impact: California's proposed supervised provisional license program. AccessLex Institute Research Paper.