What Occupational Licensing Requirements Protect the Public? Evidence from the Legal Profession

Kyle Rozema

Follow this and additional works at: https://arc.accesslex.org/grantee

Part of the Legal Profession Commons
I investigate the types of occupational licensing requirements that protect the public. To do so, I employ professional discipline as a measure of potential harm and exploit considerable state-level variation in distinctive licensing requirements for American lawyers. Using novel data from 34 states between 1984 and 2019, I find evidence suggesting that the only requirements that reduce harm are those that restrict entry for certain high-risk individuals. Even with these requirements, however, it takes over a decade following licensing for any noticeable reduction in harm to materialize, and the cumulative impact on harm reduction is small in absolute terms.

*JEL Codes: J44, D18, J21, L43, K23*

*Keywords: Occupational Licensing, Public Protection, Professional Discipline, Legal Profession, Bar Exam*
1 Introduction

Governments regulate many labor markets by requiring someone to hold a license to work in the profession. Obtaining a license often requires having certain educational qualifications and meeting certain examination requirements. For example, most American states require lawyers to earn a law degree and pass a bar exam. Occupational licensing comes at a cost to society in the form of fewer workers and potentially higher wages. The impact on the workforce size can be meaningful. For example, over the three decades between 1980 and 2019, 2.1 million law school graduates attempted to obtain a law license in different states, but almost one in ten were excluded from the profession because they were never able to pass the bar exam. Policymakers defend licensing by arguing that it ensures a workforce with a minimum level of competence, thus protecting the public from unnecessary risks.

Although considerable empirical evidence documents the effects of licensing on the labor market outcomes of workers, there is limited evidence documenting its effects on public protection. The primary reason for this limited evidence is the difficulty in finding settings with measurable indicators of harm and meaningful variation in licensing requirements. The evidence that does exist suggests that licensing prevents negative outcomes in only some professions (e.g., Kleiner and Kudrle, 2000; Currie and Hotz, 2004; Anderson et al., 2020; Farronato et al., 2020). Because licensing requirements differ across professions, one plausible explanation for why licensing appears to matter in only some professions is that specific requirements operate differently to affect the profession. For example, some requirements may exclude more high-risk workers from the profession than others, and some requirements may lead to more training than others.

In this article, I leverage a unique institutional setting to examine why only some licensing requirements reduce harm: the American legal profession. One distinctive feature of this setting lies in the nature of state bar exam policies. To pass the bar exam, a test taker must receive the state’s minimum score, known as the “cut score.” Test takers who fail a state bar exam once can retake the exam again and obtain a license if they pass on the second attempt. However, if a test taker fails a second or subsequent time, nearly half of states have limited the number of times they are allowed to retake the exam. This distinctive feature allows me to disentangle any effects of a more difficult bar exam into that driven by excluding test takers who would have passed the bar exam had the limit on the number of allowed attempts been lifted. Another
distinctive feature of this setting is that states adopt various other licensing requirements that operate only through certain causal channels, such as the required passage of an additional exam on ethics that does not, in practice, exclude individuals from the profession. This rich and idiosyncratic context enables an exploration into how particular requirements may contribute to a reduction in harm, ultimately shedding light on why only certain licensing requirements are effective in achieving this goal.

There are four primary identification challenges. First, harm inflicted on the public is not directly observable. Building on prior research that uses professional discipline as a relevant outcome in occupational licensing, I employ public disciplinary actions ordered against lawyers by state discipline bodies as a measure of lawyer quality. Disciplinary actions serve both as a source of information on the damage inflicted by lawyers and as a signal of the presence of incompetent or unethical lawyers. In an extensive data collection effort, I collected novel data on disciplinary actions in 34 states from 1984 to 2019. These data come from official state records. Importantly, these data allow me to identify not only the year lawyers were disciplined, but also the year that every disciplined lawyer obtained their license. Without information on the year that disciplined lawyers obtained a license, I would not be able to map disciplinary actions to the relevant licensing requirements when lawyers entered the profession.

Second, measuring the difficulty of passing a state bar exam is not obvious because it depends on both the cut score and the number of times test takers are allowed to retake the exam. To measure bar exam difficulty in a way that incorporates both of these policies, I use the share of test takers expected to be excluded from the profession by the bar exam. To construct this measure, I draw inspiration from a public finance literature by calibrating a structural model and simulating the exclusion rate given state bar exam policies (e.g., Currie and Gruber, 1996).

Third, different cohorts of lawyers may be subject to varying disciplinary systems and annual license renewal policies over their careers. Because changes in enforcement and renewal policies could be correlated with bar exam difficulty, failing to account for them can lead to biased estimates. To account for changes in enforcement and renewal policies, I tailor difference-in-differences research designs. In particular, bar exam difficulty applies at the state-cohort level, but I observe discipline at the state-cohort-year level, so I control for state-level changes in enforcement and renewal policies by including state-year fixed effects. This approach overcomes identifications concerns arising from changes in enforcement policies when using citation or
discipline rates as an outcome (Mungan, 2024).

Finally, the recent innovations in difference-in-differences designs that use variations of Two Way Fixed Effects estimators cannot currently accommodate state-year fixed effects which are necessary in my setting (e.g., Abraham and Sun, 2021; Callaway and Sant’Anna, 2021; Goodman-Bacon, 2021; Roth et al., 2023). To be able to exploit 53 state-level changes in the difficulty of passing the bar exam between 1984 and 2019, I therefore build on the stacked event study approach from Cengiz et al. (2019) but modify it to include two time dimensions: the year that a lawyer obtained a license which is relevant for the bar exam, and the year that a lawyer is disciplined which is relevant for enforcement policies. By modifying the research design in this way, I am able to include state-year fixed effects while maintaining the essential features of the design in Cengiz et al. (2019).

Using this research design, I estimate the effect of the difficulty of the bar exam on the quality of lawyers who obtain a license. To capture different aspects of lawyer quality, I collected data on the sanction imposed and separately study three outcomes. First, to capture the ability of licensing requirements to identify bad actors, I use the rate at which lawyers are disciplined for the first time. Second, to capture not just bad actors but also repeat offenders, I use the rate of any disciplinary action. Finally, to capture the ability of licensing requirements to identify lawyers whose conduct was determined to make them high enough risk to prevent them from being part of the profession, I use the rate at which lawyers are disbarred.

Using these measures, I find that the average change in bar exam difficulty decreases the discipline rate by 4 percent and the disbarment rate by 13 percent. I also find that these effects are robust to alternative estimation strategies and that they are primarily driven by the most difficult bar exams. However, although the estimate on the first-time discipline rate is negative, it does not reach statistical significance at conventional levels. Overall, these results suggest that even if a more difficult bar exam is not a great tool to protect against bad actors generally, it does protect against the very worst actors – those who are repeat offenders and those who would be eventually barred from practicing law.

Next, I investigate the extent to which the effects differ over lawyers’ careers. I find that the effects only emerge more than a decade after obtaining a license, and the effects are small until around two decades after obtaining a license. Additionally, these effects apply to a low baseline discipline rate: over a career, roughly one in twenty lawyers are disciplined and roughly one in fifty are disbarred. Together, these imply that the cumulative effects over the majority
of lawyers’ careers are small in absolute terms. For example, the average change in difficulty decreases the share of lawyers who are disbarred within 30 years from 1.4 to 1.2 percent.

Finally, I investigate the channels connecting licensing requirements and lawyer quality. There are various ways that licensing requirements can impact lawyer quality. For instance, a difficult bar exam may filter for high-quality lawyers, and it may lead law school graduates to study more and become more competent. I begin by estimating the extent to which excluding test takers who fail the bar exam multiple times is driving the primary results. By definition, this channel only affects cohorts of lawyers by excluding certain people from the profession. I find that the entire effect of the difficulty of the bar exam is driven by lawyers who would pass the exam but were excluded from the profession by the limit on the number of allowed attempts.

I then investigate underlying channels by estimating the impact of two other licensing policies that could impact lawyer quality in different ways. First, I estimate the extent to which state bans on individuals with felony criminal records affect lawyer quality. This policy only affects cohorts of lawyers by excluding certain people from the profession. Second, I estimate the extent to which required passage of an additional exam on ethics affects lawyer quality. This exam does not, in practice, exclude law school graduates from the profession. Therefore, this analysis sheds light on the extent to which additional training and other non-exclusion channels improve lawyer quality. Overall, I find further evidence suggesting that the relationship between licensing requirements and lawyer quality is primarily driven by the exclusion of certain high-risk lawyers from the profession.

This article makes two contributions. First, it contributes to an understanding of the economics of occupational licensing and specifically its impact on public protection. It does so by exploiting a distinctive institutional setting that can be used to explore why only some licensing requirements reduce harm. If the results are generalizable to other professions, it is possible that the varying effects of licensing requirements on public protection across different professions can be attributed to their differential impact on excluding high-risk workers from those professions. In a broader context, considering the limited empirical evidence on the effect of occupational licensing on public protection (e.g., Kleiner et al., 2016; Anderson et al., 2020; Hall et al., 2020; Larsen, 2020; Larsen et al., 2020), this article contributes to the literature by studying one of the largest and most important professions in the United States. In fact, the sheer magnitude of the potential impact of licensing in the legal profession underscores
the importance of the specific findings. Since 1980, over 200,000 law school graduates have been barred from practicing law because of their inability to pass the bar exam, and more than 300,000 other lawyers experienced a delay of at least six months in starting their careers because they failed the bar exam before ultimately passing it. It is therefore possible that the difficulty of the bar exam represents one of the most influential occupational licensing policies affecting individuals seeking entry into any profession in the United States.

Second, this article builds on a growing economics literature documenting the effects of different ways of measuring and scaling exam performance on various economic outcomes (e.g., Cawley et al., 1999; Ballou, 2009; Lang, 2010; Barlevy and Neal, 2012; Bond and Lang, 2013; Neal, 2013; Jacob and Rothstein, 2016; Nielsen, 2017). By using a parameterization of the difficulty of the bar exam that captures both the cut score and the number of allowed attempts on a single dimension, my approach introduces a way to translate standard measures of exam difficulty—usually the minimum passing score—into other measures that map onto economic outcomes better.

This article proceeds as follows. Part 2 describes the institutional setting. Part 3 describes the data and reports descriptive statistics. Part 4 describes the research design. Part 5 reports the primary results. Part 6 investigates mechanisms. Part 7 investigates the robustness of the results. Part 8 concludes.

2 Background

In every state, someone must hold a state license to practice law in that state. States justify occupational licensing as a way to protect the public from unethical and incompetent lawyers (NCBE, 2018). The legal profession has long had some of the strictest licensing requirements. Although licensing requirements for lawyers are set at the state level and thus differ from state to state, lawyers must typically obtain a three-year law degree from an accredited law school, pass a bar exam, and show that they possess good moral character by submitting to a background check.

The bar exam is a multi-day exam and is the primary barrier to becoming a lawyer (Rozema, 2023). As of 2019, every state except Wisconsin requires recent law school graduates to pass their bar exam to obtain a license (in Wisconsin, graduates of law schools located within the state do not have to pass the bar exam to practice in the state because they are granted a so-called diploma privilege). The bar exam in most states usually consists of a multiple choice
part called the Multistate Bar Exam and an essay part. Most states set the passing score based on a combination of the exam parts known as the cut score.

If someone fails a bar exam, many states allow them to retake it as many times as they would like. However, some states have set a limit on the number of attempts at some point since 1980. In some of these states, test takers can apply for special permission to retake the exam beyond the limit, but many states strictly enforce their limit and rarely grant special permission (e.g., Kansas, Rhode Island).

Lawyers can obtain a license after passing the bar exam and after the state determines they have the requisite character and fitness to practice law. Lawyers can be licensed in multiple states as long as they meet the requirements in each state. Although recent graduates usually have to take at least part of a bar exam in each state where they seek a license, some experienced lawyers seeking to obtain a license after practicing law in another state may do so without taking that state’s bar exam. However, the rules for transferring a license vary by the state someone departs from and the state they are moving to. For an experienced lawyer to obtain a license without passing the bar exam, they must typically practice law for the last 5 of 7 years.

Once a lawyer is licensed in a state, they are subject to the state’s rules of professional conduct. A disciplinary body in the state enforces the rules of conduct. State disciplinary bodies operate independently of each other, and the members of state disciplinary bodies are appointed separately by each state. The purpose of taking disciplinary action is “not to punish the attorney but, rather, to protect the public, to preserve public confidence in the legal profession, and to maintain the highest possible professional standards for attorneys” (Chadwick v. State Bar (1989) 49 Cal.3d 103, 111). Investigations of misconduct almost always start after the discipline body receives a complaint from a client or a member of the public about conduct that might constitute a violation of the professional conduct rules. Although complaints are common (e.g., roughly 1 allegation was filed for every 10 lawyers in 2017, American Bar Association, 2017), the large majority do not result in a disciplinary action (e.g., 6 percent of complaints in 2017 lead to a disciplinary action). The underlying conduct that leads to discipline includes, among other things, representing a client without diligence (e.g., missing court deadlines), representing a client despite having a conflict of interest (e.g., advising a client based on benefits the lawyer rather than to the client), participating in conduct that is prejudicial to the administration of justice (e.g., filing paperwork with the court containing
false information), and committing crimes (e.g., stealing client funds).

Discipline can be either private or public, and whether discipline is public usually depends on the sanction imposed. For example, states following the American Bar Association Model Rules make discipline private if the sanction includes, among other things, private admonishment, participation in a diversion program like going to counseling, and an advisory letter specifying a minor ethical violation; all other sanctions—disbarment, suspension, probation, or reprimand—are made public (Rule 10). In 2017, 58 percent of disciplinary actions were public. Private discipline does not lead to a public report and information about them is not a matter of public record.

I use public discipline against a lawyer as a measure of misconduct and harm. This measure filters out mere complaints against lawyers that the state disciplinary body has found to be without merit. Of course, disciplinary action does not measure the true amount of lawyer misconduct, and it does not measure the true amount of harm from the misconduct. Instead, it is an imperfect proxy for underlying misconduct and harm. Misconduct is the actual, substantive act; harm can result from the misconduct in varying degrees; and disciplinary action is the result of the potentially ensuing process. The main drawback of using public disciplinary action as a measure of misconduct and harm is that it reflects a non-random set of incidents, where several sources of variation drive the relationship between underlying unobserved misconduct and observed disciplinary action. For example, there is variation in individual lawyers’ propensity to hide misconduct, and there is likely variation in the discovery of the incident and resulting harm by victim type (e.g., Rozema, 2024). Although disciplinary action is an imperfect proxy for misconduct and harm, empirical research in other professions use similar proxies, including in medicine (e.g., Frakes, 2013; Liu and Hyman, 2019; Studdert et al., 2019), corporate management (e.g., Parsons et al., 2018), financial advising (e.g., Egan et al., 2019), mortgage lending (e.g., Piskorski et al., 2015), and policing (e.g., Dharmapala et al., 2020; Rozema and Schzenbach, 2023). The nature of selection in the context of lawyers is arguably no worse than the nature of selection in some of these other contexts.

3 Data and Descriptive Statistics

Data on Occupational Licensing Policies. The Comprehensive Guide for Bar Admission Requirements publishes state licensing policies each year. The first publication was in 1984. I use these publications to code licensing policies in states for lawyers obtaining their licenses.
from 1984 to 2019. This includes the two policies related to the difficulty of the bar exam (the cut score and the number of allowed attempts). As discussed below, in estimating the impact of a more difficult bar exam, I control for other licensing policies. Moreover, I will separately study the impact of other licensing policies as a way to investigate mechanisms. I use the Comprehensive Guide to code two other policies. First, I code whether law school graduates with a felony conviction can be admitted to practice law. These “felony bans” apply at the state-cohort level. Second, I code whether applicants are required to pass an additional exam on professional responsibility called the Multistate Professional Responsibility Examination (MPRE). Unlike the 2-3 day bar exam, the MPRE is a two-hour exam with 60 multiple-choice questions that law students typically take during their second or third year of law school. Just like the bar exam and the felony ban, the MPRE applies at the state-cohort level.

Data on Lawyer Discipline. In the United States, no centralized government-run organization compiles the disciplinary actions taken against lawyers. Instead, each state has a discipline body that manages the licensing and discipline of lawyers. All states are required by State Supreme Court rules to publish information on public disciplinary actions, but how disciplinary actions are published varies across states. I built a dataset of disciplinary actions from information on public websites, state bar magazines, and newspapers. I use official sources in each state.

For some states, I wrote automated scripts to scrape lawyer-level public web pages. For other states, hundreds of issues of state bar journals were obtained and searched in order to document the details of each public disciplinary action. Searches at the lawyer level were then conducted to identify their year of admission. Through this labor-intensive process, I built a dataset on disciplinary actions against lawyers by state discipline bodies that identifies the lawyer’s name, the year of discipline, the sanction imposed, and the year the lawyer obtained a license. Because identifying the licensing requirements when someone obtained a license requires information on the year a disciplined lawyer was licensed, I only use states where I could collect this information reliably. My sample does not contain all states because of the inability to code the year that disciplined lawyers obtain a license.

Constructing Measures of Lawyer Quality. The American Bar Association publishes information on the number of lawyers entering the profession in each state each year. I obtained historical PDFs of these publications back to 1984 and built a dataset on the number of lawyers
who obtain a license in a year after passing the bar exam from 1984 to 2019. To construct measures of lawyer quality, I divide different measures of the number of disciplinary actions from a state-cohort by the number of lawyers in the cohort.

Constructing measures of lawyer quality from disciplinary actions should capture the type of harm that a more difficult bar exam would be expected to prevent. However, it is not obvious what type of harm a more difficult bar exam would be expected to prevent. This is because the bar exam is supposed to measure lawyer competency and thus ensure a minimum competency, but competency is just one facet of how lawyers can harm the public. Some harm may stem from incompetence (e.g., missing a filing deadline), but it’s possible that the very worst outcomes arise from unethical conduct that is not necessarily linked to competence (e.g., highly competent lawyers engaging in dishonest behavior). Therefore, the bar exam may ensure a minimum competency, but it is unclear the extent to which it would translate into preventing bad outcomes generally or the very worst outcomes.

To capture different aspects of lawyer quality, I use data on the sanctions imposed to separately study three outcomes: the rate of first-time disciplinary action against a lawyer, the rate of any disciplinary action against a lawyer, and the rate that lawyers are disbarred. I separately study these three measures because they capture different aspects of lawyer quality and harm: first-time discipline identifies state recognition that a lawyer violated the rules of conduct, any discipline takes account of repeat offenders, and disbarment captures the ability of licensing requirements to filter out lawyers whose conduct was determined to make them high enough risk to prevent them from being part of the profession. Note that lawyers who violate the terms related to a first-time disciplinary action can be disciplined again for the violation, meaning they can be disciplined more than once for the same underlying conduct. Moreover, state discipline bodies are more likely to publicly discipline a lawyer with a discipline record. For these reasons, the interpretation of the rate of any discipline differs from the other measures.

To compute these rates at the state-cohort-year level, I take into account the fact that the size of the state-cohort population can change over time due to disbarments. In particular, I define the rates as the number of disciplinary actions in a year from a state-cohort divided by the number of lawyers who entered the profession who have not been disbarred as of that year. Although my primary specification accounts for disbarred lawyers in this way, the results are consistent if disbarred lawyers are not subtracted off, which is not surprising given the small
share of disbarred lawyers.

**Descriptive Statistics.** Figure 1 reports descriptive statistics of licensing policies, and Figure 2 reports the state-years that disciplinary actions are in my sample along with changes in licensing policies. The discipline data covers 36 states, including small and large states and states in different geographic regions. Across all states and years between 1984 and 2019, cut scores ranged from 125 to 145, with a mean and median cut score of 134. Between 1984 and 2019, states changed their cut score 46 times, with almost 3 out of 4 changes increasing the cut score. Between 1984 and 2019, 32 states have set a limit on the number of attempts at some point. Across all states and years since 1984, 40 percent limit the number of allowed attempts, and the mean and median number of allowed attempts conditional on having a limit is 3.5 and 3. Between 1984 and 2019, states changed their limit 33 times, with almost 3 out of 4 changes expanding the number of allowed attempts.

Figure 3 reports descriptive statistics on the size of cohorts and the annual number of disciplinary actions against lawyers. Panel A reports the distribution of the size of state-cohorts. The mean and median cohorts have 1,159 and 569 lawyers. For the smallest cohorts, 10 percent have less than 147 lawyers, and 5 percent have less than 96 lawyers. For the largest cohorts, 10 percent have more than 2,822 lawyers, and 5 percent have more than 5,500 lawyers. Panel B reports the distribution of the number of disciplinary actions at the state-cohort-year level and reveals that 46 percent have at least one first disciplinary action, 53 percent have at least one disciplinary action, and 23 percent have at least one disbarment.¹ Panel C reports the relationship between the size of the cohort and the number of disciplinary actions. It reveals a strong positive relationship between the size of the cohort and each of the measures of disciplinary actions.

Figure 4 reports descriptive statistics of disciplinary actions over lawyers’ careers, across cohorts, and over time. Panel A reports the cumulative measures of public discipline over the number of years of experience. Within 10 years, 1.2 percent of lawyers have been disciplined and 0.3 percent have been disbarred. Within 20 years, 3.1 percent of lawyers have been disciplined and 0.9 percent have been disbarred. Within 30 years, 4.8 percent of lawyers have been disciplined and 1.4 percent have been disbarred. The overall discipline rate per 100 lawyers that includes lawyers’ first and any subsequent disciplinary action is 1.5 within 10 years, 4.2

¹Figure A1 in the Appendix reports the distribution of the discipline rates.
within 20 years, and 6.9 within 30 years.

Panel B reports the evolution of the cumulative discipline rate for each cohort. Apart from the five-year period after lawyers obtained their licenses when lawyers receive few disciplinary actions, it reveals that the time paths of lawyers who start their careers in different years follow a roughly linear path. The fact that the lines rarely cross provides some evidence that different cohorts of lawyers have followed similar discipline trajectories. Panel C reports different measures of lifetime discipline by cohort. The observed downward trend in the measures is expected because lawyers who have been licensed for a longer period have had more exposure to be disciplined. The figure reveals a roughly linear trend for each of the measures, providing further evidence that overall annual discipline rates are not meaningfully different across cohorts. Finally, Panel D reports distributions of the lifetime discipline rate across states, separately by cohort. It reveals considerable between-state variation in lifetime discipline within a cohort. For example, the state at the 90th percentile witnesses a 132 percent higher rate than the state at the median for the same cohort. This variation could be explained by either true underlying differences in the misconduct propensity of lawyers in different states or by differences in the enforcement of misconduct within a state. This highlights the importance of using within-state-year variation in discipline to estimate the relationship between licensing requirements and lawyer quality.

4 Research Design

This section sets out a research design to estimate the effect of licensing requirements on lawyer quality. Section 4.1 develops a measure of the difficulty of passing state bar exams. Section 4.2 describes my identification strategy to overcome challenges from using the discipline rate as an outcome measure.

4.1 Measuring the Difficulty of the Bar Exam

Economists have only relatively recently begun understanding how the choice of using exam scores to measure ability or competency biases estimates. As Jacob and Rothstein (2016) note, the “scores that [modern exams] produce are generally not unbiased measures of student ability, and may not be suitable for many secondary analyses that economists would like to perform.” In my context, the cut score is one measure of minimum competency because test takers must obtain the score to pass the bar exam, but it is only one policy influencing the difficulty of passing a state bar exam. This is because the difficulty of passing a state bar exam
depends on the cut score of a single exam and the number of times test takers are allowed to retake the exam. The way these policies are measured does not allow them to be translated into an obvious single measure that captures the difficulty of passing the bar exam. One reason is that the expected share of the lawyer population that is affected by a change in the bar exam can differ depending on the baseline cut score and the number of allowed attempts. For example, given the distribution of bar exam scores, there are fewer test takers with scores of 130 or 131 than with scores of 145 or 146, so a 1 unit increase in the cut score from 130 would likely lead to lower additional share excluded than a 1 unit increase in the cut score from 145; and increasing a cut score from 130 to 135 will exclude more lawyers in a state that allows graduates to retake the exam three times compared to another state that allows graduates to retake the exam as many times as they would like.

These examples highlight how the cut score and the number of allowed attempts do not map onto the share of test takers who are influenced by bar exam policies. The examples also highlight how a natural way to summarize the difficulty of passing the bar exam is in terms of the share of test takers who never pass the exam. For example, if a change in the difficulty of a bar exam from either a change in the cut score or the number of allowed attempts leads the exclusion rate to increase from 10 to 15 percent, a natural way to summarize that change in difficulty is that it impacts 5 percent of the test taker population.

Although this “exclusion rate” has these nice features, using the actual exclusion rate to estimate the relationship between bar exam difficulty and discipline could lead to biased estimates because economic and other conditions might be correlated with both bar exam passage and misconduct propensity. For example, a state recession could be associated with bar exam passage rates (such as from the stress of not having a job) and a permanent impact on misconduct propensity (such as from not gaining legal experience right after law school graduation). As a result, using the actual exclusion rate as a source of variation in exam difficulty could cause a spurious correlation between bar exam difficulty and discipline.

To construct a measure of bar exam difficulty that does not suffer from these concerns while simultaneously placing changes in the cut score and number of allowed attempts on a single dimension, I draw inspiration from a public finance literature facing a similar problem in characterizing the scope of Medicaid expansion (e.g., Currie and Gruber, 1996; Cutler and Gruber, 1996). My approach is to construct a measure of bar exam difficulty that only varies with a state’s bar exam and not with the characteristics and behavior of test takers or economic
conditions. Specifically, I model test taker responses using a three-parameter Item Response Theory (IRT) model (e.g., Ballou, 2009). I do so because it is the structure that is assumed in scoring the Multistate Bar Exam that is used on the bar exam in almost all states (Albanses, 2015). As a result, it serves as a good model for test taker responses because it would only be properly used by the test administrator if the structure fits the data.

Let $\epsilon_{ij}$ be a mean-zero iid random variable that affects whether test taker $i$ answers question $j$ correctly. Let test taker competence be $\theta_i$, and let $p_{ij} = \text{pr}[f(\theta_i, j) > \epsilon_{ij}]$ be the probability that test taker $i$ answers question $j$ correctly. For test takers with the same competency ($\theta_i$), the value that the function $f(\cdot)$ takes for a question ($j$) is the same, but the (logistic) random competent $\epsilon_{ij}$ allows test takers with the same competency to answer the same question differently. The three-parameter IRT model has a structure for $f(\cdot)$ given by Equation 1.

$$f(\theta_i, b_j, a_j, c_j) = c_j + (1 - c_j) \frac{1}{1 + e^{-a_j(\theta_i - b_j)}}$$

(1)

where $b_j$ is referred to as the difficulty parameter, $a_j$ is referred to as the discrimination parameter, and $c_j$ is referred to as the guessing parameter.\(^2\) I calibrate the model to the 2019 national test taker score distribution, a representative set of question-level parameters that is used as an example by the test administrator, and several other observed moments for the rate that test takers who fail the bar exam repeat the exam and pass on subsequent attempts (see Appendix B for a full description of the model calibration). I then use the calibrated model to simulate the exclusion rate for each pair of cut scores and the number of allowed attempts.

Notably, because the simulated exclusion rate is based on a nationally representative population, it does not capture differences in lawyers’ competency across states. To the extent that test taker competency differs across states, differences in the simulated exclusion rate across states will not necessarily map onto differences in the actual exclusion rate across states. This is a feature and not a limitation of the measure. Specifically, by characterizing the difficulty of the exam as distinct from how it played out in the state, I purge it of all the effects of other factors that may be correlated with both bar exam passage rates and discipline. It is worth

\(^2\) Panel A of Figure B1 in the Appendix illustrates the interpretation of each parameter by plotting two example Item Response Probability Curves. As shown in the figure, the difficulty parameter is the test taker competency at the midpoint between the probability of the lowest competency test taker answering a question correctly and 1, the discrimination parameter is the slope of the curve at the midpoint between the probability of the lowest competency test taker answering a question correctly and 1, and the guessing parameter is the probability of the lowest competency test taker answering a question correctly.
noting that although it’s possible that the difficulty of the state bar exam can change without
a change in either the cut score or number of allowed attempts, bar exam administrators take
considerable effort to maintain a constant difficulty if the exam’s scope is changed to cover
more topics, different legal topics, or different types of questions (e.g., Albanses, 2015).

Figure 5 reports descriptive statistics on the simulated exclusion rates. Panel A reports
the simulated exclusion rates for different values of cut scores and the number of allowed
attempts. It reveals that moving to or from 3 allowed attempts drives the largest differences in
the simulated exclusion rate. For example, if a state has the average cut score of 134, moving
from 4 allowed attempts to 3 allowed attempts is associated with a 3.1 percentage point increase
in the simulated exclusion rate; in comparison, if a state has 4 allowed attempts, moving from
the lowest cut score of 125 to the highest cut score of 145 is associated with a 0.9 percentage
point increase in the simulated exclusion rate.

Panel B reports the average simulated exclusion rate for states in my sample between
1980 and 2019. It reveals that states with the most difficult bar exams in terms of the share of
lawyers expected to be excluded from the profession are not necessarily those with the highest
cut scores. For example, California is known to have a difficult bar exam because of its high
cut score, but states that meaningfully restrict the number of allowed attempts are actually
more difficult to pass in terms of the share of test takers excluded from the profession.

4.2 Difference-in-Differences

The relevant unit for which the difficulty of a bar exam applies is the state-cohort. The reason
is that bar exam difficulty is only relevant at the initial licensing stage and is not
retroactive to lawyers who already have a license. For example, if a state increases its cut score,
lawyers who were licensed before the increase do not have to retake the bar exam even if they
earned a score below the new cut score. Although the relevant unit for which the difficulty of
a bar exam applies is the state-cohort, other policies apply at the state-year level. Some of
these other policies can influence both underlying misconduct and observed discipline, creating
unique identification challenges. To formalize the identification challenges, let the discipline
rate of a state-cohort in a year be given by Equation 2.
\[ d_{cst} = r_{st}(m_{cst}) \]  

where 

\[ m_{cst} = m_{cs} + \text{Bar Exam}_{sc} + a_{t-c} + \text{Licensing Policy}_{sc} + \text{Renewal Policy}_{st} \]  

for cohort \( c \), state \( s \), and year \( t \). In Equation 2, \( d_{cst} \) is the observed state-cohort-year discipline rate, \( m_{cst} \) is the unobserved misconduct rate, and \( r_{st}(\cdot) \) is the unobserved enforcement technology for discovering, investigating, and imposing public discipline in a state-year. In Equation 3, the misconduct rate \( m_{cst} \) is at the cohort-state-year level and is a function of the baseline misconduct rate for a state-cohort (\( m_{cs} \)) and changes in this baseline that are attributable the difficulty of the bar exam (\( \text{Bar Exam}_{sc} \)), experience (\( a_{t-c} \)), a set of other initial licensing policies (\( \text{Licensing Policy}_{sc} \)) (e.g., required passage of an exam on professional responsibility), and a set of requirements to renew a law license in a given year (\( \text{Renewal Policy}_{st} \)) (e.g., continuing legal education). The enforcement technology, other initial licensing policies, and renewal policies could bias estimates on the relationship between bar exam difficulty and discipline if changes in bar exam difficulty coincide with changes in them.

To overcome these concerns, I use a modified version of the stacked event study approach from Cengiz et al. (2019). I use a stacked event study approach instead of a Two Way Fixed Effects (TWFE) approach or a synthetic difference-in-differences approach for three reasons. First, although I report estimates from a modified version of a TWFE approach and find consistent evidence, there are limitations associated with standard TWFE approaches in settings like this with variation in the timing of policy changes (e.g., Goodman-Bacon, 2021). Second, although recent advancements in TWFE approaches like Callaway and Sant’Anna (2021) address concerns in standard TWFE approaches like Callaway and Sant’Anna (2021) address concerns in standard TWFE approaches, they were not designed to accommodate potential changes in enforcement technology like in my context, and I am unaware of ways to tailor them to do so. For example, the software implementing almost all of these advancements in TWFE estimators simply does not run if, as in my setting, there are multiple observations for each unit and period (where the period triggering treatments in my context is the cohort and where I have multiple years for each cohort).  

Third, the current implementation of state-of-the-art synthetic difference-in-differences from Arkhangelsky et al. (2021) requires a binary treatment and at least some units that are not exposed to the treatment, but my setting does

---

\(^3\)For a great repository of the available software to implement recent developments in the difference-in-differences research designs, see Naqvi (2023).
not have either of these features. In particular, the treatment—difficulty of the bar exam—is continuous, and all states have a bar exam.

I therefore follow the stacked event study approach from Cengiz et al. (2019) but with one modification. An event \( e \) is a change in the difficulty of a state bar exam. Because the treatment applies at the state-cohort level, event time \( j \) identifies cohort-relative-to-the-event. For the event window, I follow Cengiz et al. (2019), who also use state-level variation, and use an 8-year event window consisting of 3 cohorts before the event and 5 cohorts after the event. As in Cengiz et al. (2019), each event has a control group consisting of other states. For each event, I generate a separate dataset that is a balanced panel of cohort-state-years consisting of a treatment state and control states. I exclude events if another event occurred in the treated state in the event window, and I exclude states from the control group if it changed its bar exam difficulty in the event window. After these restrictions, there are 38 events, consisting of 21 changes in the cut score and 17 changes in the number of allowed attempts.

Compared to Cengiz et al. (2019), the one difference is that each event-event time has multiple observations consisting of years \( t \) after cohort \( c \) obtains a law license. Using this setup, I estimate Equation 4.

\[
d_{scetj} = \alpha + \beta \text{Bar Exam}_{scej} + \gamma \text{Licensing Policy}_{scej} + \kappa_{t-c} + \psi_{es} + \phi_{ec} + \delta_{st} + \varepsilon_{scetj} \tag{4}
\]

where \( \kappa_{t-c} \) are experience fixed effects, \( \psi_{es} \) are event-by-state fixed effects, \( \phi_{ec} \) are event-by-cohort fixed effects, and \( \delta_{st} \) are state-year fixed effects. There are four benefits of this approach. First, by identifying other policies that apply at initial licensing and controlling for them (see Figure 2), I remove potential bias from Licensing Policy \( s_c \) from Equation 3. Second, and as Cengiz et al. (2019) point out, “by aligning events by event-time (and not calendar time), it is equivalent to a setting where the events happen all at once and are not staggered; this prevents negative weighting of some events that may occur with a staggered design (Abraham and Sun, 2021)”. Third, and as Cengiz et al. (2019) point out, “by dropping all states with any events within the 8 year event window, [it] further guard[s] against bias due to heterogeneous treatment effects.” Finally, the benefit of my modification to the approach is that it eliminates the remaining sources of bias in Equations 2 and 3. In particular, the state-year fixed effects \( \delta_{st} \) control for the unobserved enforcement technology \( r_{st}(\cdot) \) and any changes in renewal policies (Renewal Policy \( s_t \)). This is because changes in enforcement and renewal policies affect all lawyers moving forward, not just those who obtain a license after the change. Note that state-
year fixed effects are perfectly collinear with renewal policies within a state, so any controls for renewal policies drop out of the regression.

Following Cengiz et al. (2019), I cluster standard errors at the event-by-state level. There are three primary identifying assumptions: (1) that discipline would develop similarly over time in the treated and control states, (2) that I am able to adequately control for other initial licensing policies, and (3) that the enforcement technology does not differ by the set of initial licensing policies when a lawyer obtains a law license (this was assumed away in Equation 3 because the enforcement technology \( r_{st}(\cdot) \) did not depend on cohort \( c \)). For assumption (1), I find evidence consistent with parallel pre-trends (see below). For assumption (2), to the extent that unobserved initial licensing policies are correlated with the difficulty of the bar exam, then the estimates only capture the effects of the combination of bar exam difficulty and other unobserved policy changes. That said, I am able to include arguably all of the relevant initial licensing policies as controls except for those related to internal practices for character and fitness review (which are not publicly available). For assumption (3), it is reasonable to assume that it holds in practice because states claim to apply the rules of professional conduct without regard to the set of policies in place at initial licensing.

The coefficient of interest is on the difficulty of the bar exam (\( \beta \)). It should be interpreted as an Intent-to-Treat (ITT), capturing all underlying mechanisms that influence the equilibrium change in the discipline rate. This includes mechanisms that may be present in other professions (e.g., the effects of screening for high-quality workers and the effects of studying behavior), but this also includes a mechanism that arises in the specific context of the American legal profession due to the policies on the number times people can retake the bar exam. In particular, the simulated exclusion rate assumes that states strictly follow the limit on the number of allowed attempts. But at least some states grant special permission allowing some test takers to retake the bar exam above the limit. States have their own rules and standards for granting such special permission, and there is variation in how strict states are with granting special
To the extent that special permission is granted to test takers, it is likely granted in a way that is correlated with misconduct propensity. This means that grants of special permission could itself be a mechanism explaining any relationship between bar exam difficulty and discipline. This nuance alters the interpretation of the estimates and also represents a limitation of the approach.

To interpret the estimates, note that the simulated exclusion rate is a measure of exam difficulty that is intentionally exogenous to the actual exclusion rate, so the interpretation of the coefficient itself does not map onto intuitive reference points. My primary approach in interpreting the estimates is to scale the coefficients by the change in the simulated exclusion rate for the average change in difficulty (1.08 percentage points). The reason the average change is an important reference point is that if state supreme courts took the time to change a policy, then they likely think it is meaningful. This reference point is similar to some other reference points. For example, a one standard deviation shift in the simulated exclusion rate at state-cohort level is 1.36 percentage points; and if a state allows test takers to take the bar exam no more than 4 times, moving from the lowest to the highest observed cut score (125 to 144) is associated with a 0.9 percentage point increase in the simulated exclusion rate. However, given that the simulated exclusion rate is meaningfully higher if a state has 3 allowed attempts compared to more allowed attempts (Panel A of Figure 5), some other possible reference points would meaningfully alter the interpretation of the results. For example, moving from the 25th to the 75th percentile difficulty is associated with a 0.27 percentage point change in the simulated exclusion rate, and moving from the 10th to the 90th percentile difficulty is associated with a 3.55 percentage point change in the simulated exclusion rate. Because of these differences in reference points, I will often interpret estimates in terms of multiple reference points. In particular, I reference the p25-p75 and p10-p90 differences because they reflect observed variation in policies between states.

4 For example, in Texas, Rule 11(f) states: “[t]he Applicant shall demonstrate to the Board that mitigating circumstances exist and there has been a substantial change in the degree of the Applicant’s legal learning which makes it probable that the Applicant will pass the Texas Bar Examination or the Applicant shall demonstrate to the Board that substantial changes have occurred in the Applicant’s life by reason of education, work, experience, training and/or personal circumstances which make it substantially more likely that the Applicant will pass the Texas Bar Examination; [and], [t]he Applicant shall complete or have completed such additional review courses or additional legal study as the Board may require.” And in Rhode Island, Article II, Rule 1(d) states: “No person who has failed a total of five (5) bar examinations, whether in Rhode Island or in any other combination of states, districts, or territories of the United States (including the District of Columbia), will again be permitted to take the Rhode Island Bar Examination, and no special order excepting any such person from this five (5) examination limit will be granted by this Court…”
I make several sample restrictions. First, I exclude state-years where some law school graduates benefit from a diploma privilege and therefore do not have to pass the bar exam (Mississippi before 1986, Montana before 1985, South Dakota before 1984, Wisconsin in all years, and West Virginia before 1990). I do this because the bar exam is not relevant to many of the lawyers admitted in these states in these years and because Rozema (2021) studied the impact of these four changes on lawyer quality. Second, I exclude cohorts after 2015. I do this because it is rare for lawyers to be disciplined early in their career, so changes in the difficulty of the bar exam occurring in the last 5 years of my sample are not expected to have a meaningful effect on discipline during the years in my sample. Third, I exclude Alabama from the primary analysis. The reason is that including Alabama in the sample over-inflates the significance of a single change in the bar exam on the estimates. In robustness checks below, I report results that include Alabama, and I discuss why I believe the best estimates exclude Alabama from the sample.

5 Results

Table 1 reports descriptive statistics comparing the panel dataset and the stacked event study dataset. It reveals that bar exam difficulty and each measure of lawyer quality are similar across the datasets. For example, the average simulated exclusion rate is between 9.8 and 9.9 percent in both datasets, and the average annual first-time discipline rate is between 0.139 and 0.144 percent in both datasets.

Pretrends. I begin by investigating pretrends in two ways. First, I estimate Equation 4 but replace the simulated exclusion rate with interactions between the treated group and event time variables that take the value of the change in the simulated exclusion rate for the event. The coefficients should be interpreted as the effect of a 1 percentage point increase in the simulated exclusion rate on discipline in a specific year relative to the change. Roth et al. (2023) note that a common test for pre-trends is to assess whether the estimated coefficients in such a regression are different from zero. In particular, they note that “[i]f all of the pre-treatment coefficients ... are insignificant, this is usually interpreted as a sign in favor of the validity of the design, since we cannot reject the null that parallel trends was satisfied in the pre-treatment period.” Figure 6 reports the results and reveals that the coefficients in the pre-period are small and none are statistically significant. Under this test, the figure reveals no evidence against parallel pretrends.
Second, I use the approach in de Chaisemartin and D’Haultfoeuille (2020), which is to use “placebo estimators comparing the outcome trends of switchers and non-switchers with the same period-one treatment, before switchers switch.” To implement this test, I use a panel dataset at the state-cohort level where the outcomes are the mean quality measures. Under this test, I again find no evidence against parallel pretrends for each of the three outcomes (the p-values of the joint test against parallel trends are 0.77, 0.86, and 0.91 for the three outcomes).

**Impact on Lawyer Quality.** Figure 6 also reveals initial evidence that a more difficult bar exam improves lawyer quality. The coefficients are lower in the post-period than in the pre-period, which indicates that a more difficult bar exam decreases the discipline rate. Although the individual coefficients in the post-period are not statistically significant for most periods, they are jointly significant for the disbarment rate (joint estimate of -0.023, \( p < 0.003 \)). To directly estimate the effect of a more difficult bar exam, Table 2 reports regression results estimating Equation 4. In the preferred specification in Column 3, I find that a one percentage point increase in the simulated exclusion rate decreases the discipline rate by 0.79 percentage points and the disbarment rate by 0.44 percentage points. Relative to the baseline rates, these represent decreases of 3.9 percent and 11.6 percent. Alternatively, the estimates imply that the average change in difficulty decreases the discipline rate by 4.2 percent and the disbarment rate by 12.5 percent; the estimates imply that moving from 25th to 75th percentile difficulty decreases the discipline rate by 0.6 percent and the disbarment rate by 3.1 percent; and the estimates imply that moving from 10th to 90th percentile difficulty decreases the discipline rate by 13.8 percent and the disbarment rate by 41.2 percent. The estimates on the first-time discipline rate are negative and small but not statistically significant at conventional levels. Section 7 investigates the robustness of these findings in five ways, and the results of the robustness checks support the validity of the primary estimates.

**Distribution of Effects.** In Figure 6 and Table 2, the simulated exclusion rate enters linearly. To understand the extent to which the results are driven by certain parts of the distribution of bar exam difficulty, I replace the linear term for the simulated exclusion rate with indicators for the quartile of the simulated exclusion rate. In particular, I estimate separate regressions for each quartile in the preferred specification in Column 3 of Table 2. The estimated coefficient is the quartile relative to the average in the other quartiles. Figure 7 reports estimated coefficients divided by the mean of the outcome. The figure reveals that the esti-
mate on the top quartile of difficulty is lower than all other quartile estimates for all outcomes. However, apart from a higher estimate on the first quartile for disbarment, there appear to be no meaningful systematic differences in the estimates on the first, second, and third quartiles. Overall, these results provide evidence that the overall results are primarily driven by states with the most difficult bar exams.

Timing of Effects. I next assess the extent to which the effects of a more difficult bar exam differ over lawyers’ careers. I do so for two reasons. First, it is important to understand whether the impacts of a more difficult bar exam kick in right away or only in the distant future. Given that lawyers rarely get disciplined at the beginning of their careers (Figure 4, Panel A), one may expect the impact of the bar exam to have a smaller effect at the beginning of lawyers’ careers. Second, the main results could be partly driven by differences in career length of marginally excluded lawyers. For example, if a bar exam is made easier and the marginal group who obtains a license has a less successful legal practice, then they may practice for fewer years and would not be at risk of being disciplined for the same number of years. If so, the main estimates would be the lower bound of the true effect. On the other hand, if the marginal group works longer than the group that would always pass, then my estimates would be the upper bound of the true effect. Although I cannot test for these mechanisms directly, documenting whether there are differential impacts on the effect of a more difficult bar exam over lawyers’ careers can provide at least some suggestive evidence on whether marginally excluded lawyers have different career lengths. For example, if the effect of a more difficult bar exam has a non-linearity at the end of the lawyer’s careers, it would be suggestive that it is driven by differences in career lengths. In any case, if the only goal of a more difficult bar exam is to protect the public, then policymakers should be indifferent to whether the main results are driven by a lower propensity of misconduct or shorter careers.

To assess the timing of the effect of a more difficult bar exam over lawyers’ careers, I estimate a version of Equation 4 interacting the simulated exclusion rate with indicators for 5-year experience bins. I then use these estimated marginal effects for each experience bin to compute the cumulative effect as of different years of experience. In particular, I estimate the implied cumulative discipline rate and compare it to the cumulative discipline rate of the average cohort. To estimate the implied rate, I assume that the marginal effects apply to each year in the relevant experience range, and I multiply the marginal effects by the difference
in simulated exclusion rate for each of the relevant comparisons (e.g., the average change in bar exam difficulty). I acknowledge that this approach has several limitations. For example, although the marginal effects have confidence intervals, the implied cumulative effects do not. Moreover, it assumes that the relationship between the effect of a more difficult bar exam and discipline over experience is captured by the interactions between the 5-year experience bins and a linear term for the simulated exclusion rate.

Figure 8 reports the results. For the marginal effects, I find that the negative effect of a more difficult bar exam is typically close to zero and statistically insignificant in the first 10 years of lawyers’ careers but is larger and statistically significant after that. This timing is consistent with evidence on how eliminating the bar exam affects lawyer quality from Rozema (2021). For the cumulative effects, I find that the cumulative discipline rate would be almost identical for all outcomes within ten years of licensing. After year ten, the outcomes begin to diverge but the differences are still small in both absolute and relative terms. At year twenty, the average change in difficulty decreases the cumulative discipline rate from 4.2 to 4.1 percent; moving from the 25th to 75th percentile difficulty does not decrease the cumulative discipline rate; and moving from the 10th to 90th percentile difficulty decreases the cumulative discipline rate from 4.2 to 3.8 percent. At year thirty, the average change in difficulty decreases the cumulative discipline rate from 6.9 to 6.6 percent; moving from the 25th to 75th percentile difficulty decreases the cumulative discipline rate from 6.9 to 6.8 percent; and moving from the 10th to 90th percentile difficulty decreases the cumulative discipline rate from 6.9 to 5.7 percent.

6 Mechanisms

The difficulty of the bar exam can affect lawyer misconduct through a number of causal channels. To highlight how these channels could operate, Appendix C presents a simple model of test taker ability, studying behavior, bar exam scores, and misconduct. Although the model is only one way to formalize the channels, it highlights three in particular. First, bar exam difficulty can affect misconduct by excluding test takers with different competencies. If this is the only channel at work, a more difficult bar exam may be expected to decrease misconduct if exam scores are negatively correlated with misconduct. Second, bar exam difficulty can affect misconduct by affecting the distribution of test takers with different competencies in the population. For example, test takers may respond to a more difficult exam by opting out of
taking the exam, such as by taking the bar exam in another state. If this is the only channel at work, a more difficult bar exam that drives lower competency test takers away from the state may be expected to decrease misconduct.

Third, bar exam difficulty can affect the underlying misconduct rate for lawyers who obtain a license. This can operate through several specific channels that may work to increase or decrease misconduct. For example, exam difficulty can affect how much test takers prepare for the exam. If the only channel at work is to increase how much test takers know about the law through greater preparation for the bar exam, a more difficult bar exam may be expected to decrease misconduct. On the other hand, exam difficulty can affect the share and type of test takers who fail before passing, which may, in turn, put these lawyers in a worse financial situation. If the only channel at work is to increase the number of lawyers who fail the bar exam before passing it, a more difficult bar exam may be expected to increase misconduct. At least in the context of the American legal profession, the potential for licensing to backfire in this specific way is meaningful. For example, of the 1.9 million graduates who passed the bar exam between 1980 and 2019, almost one in six of them did not pass the exam on their first attempt (Rozema, 2023).

I investigate mechanisms in three ways. I begin by estimating the extent to which the primary results are driven by the exclusion of test takers who would have passed the bar exam had the limit on the number of allowed attempts been lifted. I then investigate mechanisms by assessing the impact of two other licensing policies that are separate from the bar exam.

**Bar Exam Exclusion Channel.** First, I investigate the extent to which excluding test takers who fail multiple times before passing drives the results. To identify the portion of the simulated exclusion rate that captures the test takers who would have passed the exam if the limit on the number of allowed attempts was lifted, I make two simplifying assumptions. First, I assume test takers who fail the bar exam study for the exam again and improve their scores after failing according to Equation 5.

\[
P(\theta, \alpha) = 1 - \lambda^\alpha [1 - P(\theta, 0)]
\]

where \( P(\theta, \alpha) \) is the probability of test taker with competence \( \theta \) passing on their \( \alpha \)th attempt and \( \lambda \) captures the decrease in the probability of failing on subsequent attempts. This approach assumes diminishing marginal behavioral effects. Second, I assume that every time a test taker fails the exam, they quit with probability \( \delta \), where quitting means never attempting to retake
the bar exam. Thus, the probability that a test taker retakes the exam after $\alpha$ failed attempts is $(1 - \delta)^\alpha$. If a state has a limit on the number of allowed attempts of $A$, this limit decreases the number of repeat test takers according to Equation 6.

$$\int_{-\infty}^{+\infty} f(\theta) \sum_{j=A}^{\infty} f_{i-j}(1 - \delta)^j \lambda^{j-1} (1 - P(\theta))^j$$

where $f_{i-j}$ is the number of first-time test takers in year $t$. This expression represents the number of people excluded from retaking the bar exam. Of this group, the limit excludes the test takers who would have passed from the profession. I estimate $\delta$ from descriptive statistics and calibrate $\lambda$ by moment matching (see Appendix D for details). Incorporating both parameters into the simulations, I identify simulated test takers who were excluded from the profession but who would have passed the bar exam before quitting if there was no limit on the number of allowed attempts. Based on Equation 6, these test takers are those excluded because of the limit on the number of allowed attempts, and the remaining test takers were excluded from the profession because of the cut score. Using this measure, I estimate the primary specifications but replace the single simulated exclusion rate with this portion and the remaining portion capturing everyone else.

Panel A of Table 3 reports the results using the stacked event study approach, and Panel A of Table 4 reports the results using a modified TWFE approach in Equation 7.

$$d_{sct} = \alpha + \beta X_{sc} + \sigma_c + \kappa_{t-c} + \delta_{st} + \epsilon_{sct}$$

where $X_{sc}$ is the set of licensing policies, $\sigma_c$ are cohort fixed effects, $\kappa_{t-c}$ are experience fixed effects, and $\delta_{st}$ are state-year fixed effects. I cluster standard errors by state in this approach. Note that Equation 7 is not a classic TWFE design because there are multiple observations for each state-cohort.

I find that the entire effect of the difficulty of the bar exam is driven by lawyers who would pass the exam but were excluded from the profession by the limit on the number of allowed attempts. This means that a more difficult bar exam likely works to decrease discipline by excluding certain test takers from the profession rather than improving the competency of test takers who pass the exam. This is consistent with descriptive evidence suggesting that lawyers who fail the bar exam at least once but eventually pass are at higher risk than those who pass on their first attempt (Kinsler, 2017; Anderson and Muller, 2019).
**Felony Ban.** Next, I estimate whether a policy preventing people with a felony conviction from entering the legal profession affects lawyer quality (note that this was included as a control variable but not reported in the primary results). The felony ban could only act through exclusion because it, by definition, only affects a certain group of potential test takers and does not have an obvious channel to affect the behavior of lawyers without a felony conviction.

To estimate the impact of the felony ban, I begin by using the same stacked event study design described above but define the event as a change in the felony ban. Because the stacked event study design drops events if another event occurs within the event window, it decreases the number of felony ban events from 17 to 12. Given the relatively small amount of policy variation after these exclusions, I also report results the modified TWFE specification.

Panel B of Tables 3 and 4 report the results. I find some evidence that a felony ban decreases the discipline rate. The TWFE results are statistically significant for both the discipline rate and the disbarment rate. Although the stacked event study results are not statistically significant at conventional levels, it is identified off 12 events. Moreover, the policy may be expected to influence a small share of law school graduates, so it may be difficult to detect a relationship even if one exists. Overall, this provides at least some additional suggestive evidence that licensing operates by excluding certain people from the profession.

**Ethics Exam.** Next, I estimate whether adopting a requirement to pass an exam on professional responsibility (MPRE) affects lawyer quality (note that this was included as a control variable but not reported in the primary results). In theory, the 60-question MPRE should not act through an exclusion channel because it is easier to pass than the bar exam. This means that no one who fails the MPRE would likely pass the bar exam, implying that it does not itself impact who obtains a license. It therefore could influence discipline only through training or other non-exclusion channels. The assumption that the MPRE does not exclude people from the profession is supported by empirical evidence. In particular, Rozema (2023) notes that the MPRE is required before someone can take the bar exam in most states, so if the adoption of the MPRE excludes people from the profession, then it would do so before people take the bar exam. As a way to estimate the impact of the MPRE on excluding graduates from the profession, Rozema (2023) therefore estimates whether state adoption of the MPRE affects the number of first-time test-takers and finds no evidence that it does.

To estimate the impact of the MPRE requirement, I follow both approaches that I used
to study the felony ban. The stacked event study approach uses 16 MPRE adoption events, and
the TWFE approach uses 23 adoption events. Panel C of Tables 3 and 4 report the results. I
find no evidence that the MPRE affects the discipline rate. In fact, the estimates rule out large
negative effects, suggesting that this non-exclusion channel does not improve lawyer quality.

7 Robustness

I investigate the robustness of the primary results in five ways. First, I investigate the
extent to which my modification to the stacked event study approach in Cengiz et al. (2019)
affects the results. Recall that I modified the approach out of concerns that states change
their enforcement of misconduct at the same time they change the difficulty of the bar exam. I
therefore modified the approach by introducing both cohorts and years and including state-year
fixed effects. To directly implement the approach in Cengiz et al. (2019), I compute the mean
quality measures for each state-cohort, and I estimate Equation 8.

\[ d_{scej} = \alpha + \beta \text{ Bar Exam}_{sc ej} + \gamma \text{ Licensing Policy}_{sc} + \psi_{se} + \phi_{ec} + \varepsilon_{sc ej} \] (8)

where \( \psi_{se} \) are event-by-state fixed effects and \( \phi_{ec} \) are event-by-cohort fixed effects. Because
there is only one observation for each state-cohort-event, I can no longer include experience and
state-year fixed effects. Table A1 in the Appendix reports the results reveals consistent results,
suggesting that my modifications to the approach in Cengiz et al. (2019) do not meaningfully
impact the results.

Second, I use randomization inference by performing placebo tests. To do so, I drop
treated states from the stacked event study dataset, randomly reassign one of the control states
to be the treated state in each event sub-sample, and re-estimate the primary specification. To
be able to estimate the change in lawyer quality that would be witnessed for a placebo group
that matches the observed policy variation, I add the change in the simulated exclusion rate
for the actual treated state for a given event to the simulated exclusion rate of the control state
that receives the placebo treatment in the post-period. For example, if the treated state in
an event increased its cut score in a way that leads to an increase in the simulated exclusion
rate of 0.5 percentage points, and if the control state randomly chosen to receive the placebo
treatment had a simulated exclusion rate of 8.5 percent, then the control state receiving the
placebo treatment would continue to have a simulated exclusion rate of 8.5 percent in the
pre-period but would now have a simulated exclusion rate of 9.0 percent in the post-period. I
simulate this approach 5,000 times for each outcome.

Figure A2 in the Appendix reports the distribution of placebo estimates along with the actual estimates. The effect of the difficulty of the bar exam on lawyer quality is greater than 85.1 percent of the placebo estimates for the first-time discipline rate, 94.4 percent of the placebo estimates for the discipline rate, and 99.5 percent of the placebo estimates for the disbarment rate. Moreover, the actual estimates document a meaningfully larger effect for each of the outcomes than in the placebo estimates: for the first-time discipline rate, the mean placebo estimate is -0.00005, compared to the actual estimate of -0.0034; for the discipline rate, the mean placebo estimate is 0.00007, compared to the actual estimate of -0.0079; and for the disbarment rate, the mean placebo estimate is effectively zero, compared to the actual estimate of -0.0044. Differencing out the placebo estimates from the actual estimates, this means that the actual estimates are -0.0033 larger than the mean placebo estimates for the first-time discipline rate, -0.0079 larger than the mean placebo estimates for the discipline rate, and -0.0044 larger than the mean placebo estimates for the disbarment rate. These differences are close to the range of estimated coefficients from Table 2.

Third, to understand how much single events contribute to the overall estimates, I estimate the primary specification leaving out one event at a time. Figure A3 in the Appendix reports the results and reveals that the estimates are not sensitive to the exclusion of events. In terms of the magnitude of the estimates, the coefficients for the discipline rate are between -0.0062 and -0.0092, and the coefficients for the disbarment rate are between -0.0038 and -0.005. In terms of the precision of the estimates, all but 2 of the 38 estimates for the discipline rate are significant at the 5 percent level, and all of the estimates for the disbarment rate are significant at the 5 percent level.

Fourth, to understand whether the results depend on the use of the simulated exclusion rate as a measure of bar exam difficulty, I estimate the primary specification but replace the simulated exclusion rate with various measures of the cut score and the number of allowed attempts along with interactions between them. Table A3 in the Appendix reports the results. The sign of the point estimates are consistent, but the estimates are typically much noisier, with many of the estimates not reaching statistical significance at conventional levels. This is perhaps expected given that the cut score and number of allowed attempts do not capture the economic relationship between the policies and the expected share of the profession that they impact.
Finally, I again use the panel structure of the data and estimate the modified TWFE specification in Equation 7 with different sets of controls. Table A2 in the Appendix reports the results and reveals evidence consistent with the primary results in Table 2. As discussed above, I dropped Alabama from the entire analysis, including in the TWFE specification. The reason is that Alabama is the only state that changed the number of allowed attempts to or from 2, and the resulting change in the simulated exclusion rate for this event is 6 times the change of any other change (22 percentage points compared to 4 percentage points). Given that the amount of variation from this Alabama event is orders of magnitudes greater than for other events, one may expect Alabama to meaningfully affect the estimates. This is, in fact, what I find. I re-estimate the TWFE specification while leaving out one state at a time (where I include Alabama when leaving out other states). Figure A4 in the Appendix reports the results and reveals that the inclusion of Alabama reduces the estimated size of the coefficient for all outcomes. If Alabama is included in the sample, the only estimates that are statistically significant are on the first-time discipline rate. I believe the best estimates exclude Alabama from the sample because (1) including it over-inflates the significance of a single change in bar exam difficulty, and (2) the estimates are stable to excluding all other states from the sample. Note that excluding Alabama from the sample only affects the TWFE estimates. The reason is that two events from Alabama occurred within a year of the other (see Figure 2) and I followed the stacked event study approach from Cengiz et al. (2019) and dropped events where another event occurred within the event window.

Overall, the results of the robustness checks support the validity of the primary estimates. Moreover, the results using the cut score and number of allowed attempts as a measure of bar exam difficulty highlight the value of my approach of using a simulated exclusion rate, which addresses many of the concerns summarized in Jacob and Rothstein (2016) in that failing to account for exam scores properly complicates the interpretation of estimates.

8 Conclusion

Occupational licensing is primarily justified as a means to protect the public from potential harm, but empirical evidence suggests that only some requirements in some professions achieve this goal (e.g., Kleiner and Kudrle, 2000; Currie and Hotz, 2004; Anderson et al., 2020; 5

5 Alabama had 2 allowed attempts from 1980 to 1992, 5 allowed attempts in 1993, and unlimited attempts from 1994 to 2019. Iowa has 2 allowed attempts as well, but Iowa’s policy has not changed since 1980 and therefore does not provide variation identifying the simulated exclusion rate.
Farronato et al., 2020). Because licensing requirements can differ across professions, one reason that could explain the mixed impacts is that different requirements can affect a profession—and, in turn, public protection—in different ways. This article leverages a distinctive institutional setting, the American legal profession, that allows for an examination of why some but not other licensing requirements reduce harm. By building a novel dataset on public discipline in 34 states between 1984 and 2019 and exploiting considerable state-level variation in licensing requirements, I find that the only requirements that decrease harm are those that exclude certain high-risk workers from entering the profession. However, I also find that the cumulative difference that is attributed to improvements in quality is small in absolute terms. For example, I find that over the 30 years after obtaining a license, the average change in difficulty decreases the share of lawyers who are disbarred from 1.4 to 1.2 percent.

The results inform debates about occupational licensing requirements for the American legal profession. Hadfield (2022) discusses how “studies evaluating the impact of regulation on quality in legal markets are practically non-existent,” and the only past evidence in the U.S. context is based on four policies that required graduates from four law schools to pass the bar exam for the first time (Rozema, 2021). Yet, the difficulty of state bar exams is likely the most important and policy-relevant licensing policy for the American legal profession. One reason is that, although some reformers are calling for the bar exam to be eliminated (e.g., Winston, 2023), most experts think that it is not a viable option (e.g., Hadfield, 2022). This means that changing the difficulty of state bar exams is the most policy-relevant reform, as evidenced by over 80 changes in bar exam difficulty since 1980. Another reason is that it has the largest impact on the size of the legal profession. For example, whereas eliminating the requirement to earn a law degree would increase labor supply by no more than 4 percent, replacing the strictest bar exam policies with the most lenient ones would increase labor supply by 22 percent (Rozema, 2023). By shedding light on the effects of the primary licensing policies in the legal profession, this article adds to the limited research on the effect of licensing on public protection.

Of course, this evidence is only one of the factors influencing how difficult state bar exams should be. Although I document one benefit of a more difficult bar exam, there are also costs.

---

Outside of the U.S. setting, one other study does not rely on disciplinary actions to study the relationship between licensing and lawyer quality. In particular, Ramseyer and Rasmusen (2015) study the impact of the difficulty of the Japanese bar exam on the credentials of lawyers and find a more difficult bar exam actually decreases lawyer credentials.
A more difficult bar exam meaningfully decreases the number of lawyers entering the profession (Rozema, 2023), which can in turn increase the cost of legal services and decrease access to legal services. In addition, there are concerns that a more difficult bar exam has a disparate impact on law school graduates from groups historically excluded from the legal profession (e.g., ABA, 2022), and it’s possible that a more difficult bar exam negatively impacts interstate mobility of lawyers. Given these costs, it is not obvious that states should seek to exclude some higher-risk lawyers from the profession through a more difficult bar exam. Indeed, especially given that the absolute impact of a more difficult bar exam on harm reduction is small, it is possible (or even plausible) that states would be better off allowing more lawyers into the profession with a less difficult bar exam.
References


Figure 1: Descriptive Statistics of Occupational Licensing Policies for American Lawyers

A. Bar Exam Cut Score

B. Number of Allowed Attempts on Bar Exam

C. Years with Felony Ban

D. Years with Ethics Exam

Notes: The figure reports descriptive statistics of licensing policies for the U.S. legal profession for lawyers obtaining a license between 1984 and 2019. Panel A reports the distribution of state cut scores at the state-year level. Panel B reports the number of allowed attempts at the state-year level. Panel C reports the percent of years that a state had a felony ban. Panel D reports the percent of years that a state required passage of an ethics exam called the MPRE.
Figure 2: Sample of Cohorts, Disciplinary Actions, and State Licensing Policy Changes

Notes: Each solid line indicates the state-cohorts in the sample. Each dashed line indicates state-cohorts that are not in the sample. The filled-in square markers indicate the timing of changes in the bar exam cut score. The filled-in circle markers indicate the timing of changes in the number of allowed attempts for the bar exam. The hollow diamond markers indicate the timing of changes in the rule for whether law school graduates with a felony conviction can be admitted to practice law. The × markers indicate the timing of the adoption of an ethics exam.
Figure 3: Size of Cohorts and Disciplinary Actions

A. Size of State-Cohorts

B. Number of Disciplinary Actions

C. Disciplinary Actions by Size of State-Cohort

Notes: Panel A reports the distribution of the number of lawyers at the state-cohort level. Panel B reports the distributions of three measures of disciplinary actions. Panel C reports the relationship between the size of state-cohorts and the number of disciplinary actions they receive, where state-cohort-years are split into 20 groups based on the number of lawyers in the state-cohort.
Figure 4: Discipline Over Career, Between Cohorts, and Across States

A. Over Career

B. Evolution of Discipline

C. Lifetime Discipline

D. Distribution Across States Over Time

Notes: Panel A reports the cumulative discipline rate over lawyers' careers. Panel B reports the cumulative discipline rate over time by cohort, where each line represents a cohort and lines are colored for cohorts starting their career in different decades. Panel C reports the lifetime discipline rate of different cohorts. Panel D reports the state-cohort lifetime discipline rate through a letter-value plot, where a distribution ranges from the 5th to the 95th percentile in steps of 10 percentiles.
Figure 5: Simulated Exclusion Rate as Measure of the Difficulty of State Bar Exams

A. Simulated Exclusion Rate by Cut Score and Allowed Attempts

B. Average Simulated Exclusion Rate By State

Notes: The figure reports descriptive statistics of a measure of the difficulty of state bar exams capturing the share of the national test taker population that is excluded from the profession for a given cut score and number of allowed attempts. Panel A reports this simulated exclusion rate for different cut scores, separately by the number of allowed attempts. Panel B reports the average simulated exclusion rate for each state in my sample from 1984 to 2019.
Table 1: Descriptive Statistics of Panel and Stacked Event Study Datasets

<table>
<thead>
<tr>
<th></th>
<th>Dataset</th>
<th>Panel</th>
<th>Event Study</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>A. Sample</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>States</td>
<td>34</td>
<td>34</td>
<td></td>
</tr>
<tr>
<td>Cohorts</td>
<td>37</td>
<td>37</td>
<td></td>
</tr>
<tr>
<td>Years</td>
<td>40</td>
<td>40</td>
<td></td>
</tr>
<tr>
<td>State-Cohorts</td>
<td>1,209</td>
<td>1,127</td>
<td></td>
</tr>
<tr>
<td>State-Cohort-Years</td>
<td>26,348</td>
<td>23,599</td>
<td></td>
</tr>
<tr>
<td>Observations after Restrictions</td>
<td>26,314</td>
<td>157,363</td>
<td></td>
</tr>
<tr>
<td><strong>B. Bar Exam Difficulty</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cut Score</td>
<td>133.7</td>
<td>134.0</td>
<td></td>
</tr>
<tr>
<td>Percent with Limit on Attempts (x100)</td>
<td>38.2</td>
<td>35.1</td>
<td></td>
</tr>
<tr>
<td>Simulated Exclusion Rate (x100)</td>
<td>9.8</td>
<td>9.9</td>
<td></td>
</tr>
<tr>
<td><strong>C. Lawyer Quality</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>First-Time Discipline Rate (x100)</td>
<td>0.139</td>
<td>0.144</td>
<td></td>
</tr>
<tr>
<td>Discipline Rate (x100)</td>
<td>0.199</td>
<td>0.203</td>
<td></td>
</tr>
<tr>
<td>Disbarment Rate (x100)</td>
<td>0.039</td>
<td>0.038</td>
<td></td>
</tr>
</tbody>
</table>

Notes: The table reports descriptive statistics of the panel dataset (Column 1) and the stacked event study dataset (Column 2). Panel A reports descriptive statistics of the sample. Panel B reports descriptive statistics on bar exam difficulty. Panel C reports descriptive statistics on the three measures of lawyer quality.
Figure 6: Measures of Lawyer Quality around Changes in Bar Exam Difficulty

A. First-Time Discipline Rate

B. Discipline Rate

C. Disbarment Rate

Notes: The figure reports event studies of different measures of lawyer quality around changes in the difficulty of state bar exams. The figure reports coefficients and 90 percent confidence intervals from estimating Equation 4 but replacing the simulated exclusion rate with event time variables that take the value of the change in the simulated exclusion rate for the event for the treated group. Following Cengiz et al. (2019), standard errors are clustered at the event-by-state level. The panels differ by the outcome used, as indicated above the panel. All outcomes are multiplied by 100.
Table 2: Difficulty of Bar Exam and Lawyer Quality

<table>
<thead>
<tr>
<th></th>
<th>(1)</th>
<th>(2)</th>
<th>(3)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>A. First-Time Discipline Rate</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Simulated Exclusion Rate (ppt)</td>
<td>-0.0033</td>
<td>-0.0033</td>
<td>-0.0034</td>
</tr>
<tr>
<td></td>
<td>(0.0021)</td>
<td>(0.0021)</td>
<td>(0.0021)</td>
</tr>
<tr>
<td>Mean First-Time Discipline Rate</td>
<td>0.1441</td>
<td>0.1441</td>
<td>0.1441</td>
</tr>
<tr>
<td><strong>B. Discipline Rate</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Simulated Exclusion Rate (ppt)</td>
<td>-0.0074**</td>
<td>-0.0074**</td>
<td>-0.0079**</td>
</tr>
<tr>
<td></td>
<td>(0.0033)</td>
<td>(0.0032)</td>
<td>(0.0033)</td>
</tr>
<tr>
<td>Mean Discipline Rate</td>
<td>0.2033</td>
<td>0.2033</td>
<td>0.2033</td>
</tr>
<tr>
<td><strong>C. Disbarment Rate</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Simulated Exclusion Rate (ppt)</td>
<td>-0.0042***</td>
<td>-0.0042***</td>
<td>-0.0044***</td>
</tr>
<tr>
<td></td>
<td>(0.0012)</td>
<td>(0.0012)</td>
<td>(0.0012)</td>
</tr>
<tr>
<td>Mean Disbarment Rate</td>
<td>0.0378</td>
<td>0.0378</td>
<td>0.0378</td>
</tr>
</tbody>
</table>

**Covariates**

- State-Event FE ✓ ✓ ✓
- Event-Year FE ✓ ✓ ✓
- State-Year FE ✓ ✓ ✓
- Cohort and Experience FE ✓ ✓
- Other Initial Licensing Policies ✓

Observations 157,363 157,363 157,363

*Notes: The table reports regressions results estimating Equation 4. The panels differ by the outcome: first-time discipline rate (Panel A), discipline rate (Panel B), and disbarment rate (Panel C). All outcomes are multiplied by 100. Following Cengiz et al. (2019), standard errors are clustered at the event-by-state level. * p<0.10, ** p<0.05, *** p<0.01.*
Figure 7: Shape of Relationship between Bar Exam Difficulty and Lawyer Quality

A. First-Time Discipline Rate

B. Discipline Rate

C. Disbarment Rate

Notes: The figure reports estimated coefficients for indicator variables for the quartile of the simulated exclusion rate. The specification replaces the linear term for the simulated exclusion rate in Column 3 of Table 2 with indicators for the quartile. Within a given panel, I estimate separate regressions for each quartile, so the estimated coefficient is the quartile relative to the average in the other quartiles. To compare across panels, each estimate is divided by the mean of the outcome. Coefficients and 90 percent confidence intervals are reported. The panels differ by the outcome, as indicated above the panel. All outcomes are multiplied by 100.
Figure 8: Marginal and Cumulative Effect of Bar Exam Difficulty Over Lawyers’ Career

A. Marginal First-Time Discipline Rate

B. Cumulative First-Time Discipline Rate

C. Marginal Discipline Rate

D. Cumulative Discipline Rate

E. Marginal Disbarment Rate

F. Cumulative Disbarment Rate

Notes: The left panels report the marginal effect of the simulated exclusion rate at different times in lawyers’ careers from a version of Equation 4 where I interact the simulated exclusion rate with indicators for 5-year experience bins (using the stacked event study dataset and Column 3 of Table 2). Coefficients and 90 percent confidence intervals are reported. The right panels report the implied cumulative effect of the simulated exclusion rate over lawyers’ careers estimated using the marginal effects. The blue solid line is the cumulative discipline rate of the average cohort. The long-dashed green line is the cumulative effect evaluated at the average change in bar exam difficulty. The dashed gold line is the cumulative effect evaluated at the difference between the 25th and 75th percentile difficulty. The short-dashed orange line is the cumulative effect evaluated at the difference between the 10th and 90th percentile difficulty. The panels differ by the outcome used, as indicated above the panel. See the text for further details on how the implied cumulative effects are estimated.
Table 3: Channels Connecting Licensing and Lawyer Quality

<table>
<thead>
<tr>
<th></th>
<th>Outcome: Discipline Rate</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>First (1)</td>
</tr>
<tr>
<td>A. Bar Exam Difficulty Channels</td>
<td></td>
</tr>
<tr>
<td>Simulated Exclusion Rate from</td>
<td>-0.0034</td>
</tr>
<tr>
<td>Number of Allowed Attempts (ppt)</td>
<td>(0.0021)</td>
</tr>
<tr>
<td>Simulated Exclusion Rate from</td>
<td>0.0002</td>
</tr>
<tr>
<td>Cut Score (ppt)</td>
<td>(0.0221)</td>
</tr>
<tr>
<td>Mean Rate</td>
<td>0.1441</td>
</tr>
<tr>
<td>Number of Events</td>
<td>38</td>
</tr>
<tr>
<td>Observations</td>
<td>157,363</td>
</tr>
<tr>
<td>B. Felony Ban</td>
<td></td>
</tr>
<tr>
<td>Ban on Applicants with Felonies</td>
<td>0.0056</td>
</tr>
<tr>
<td></td>
<td>(0.0084)</td>
</tr>
<tr>
<td>Mean Rate</td>
<td>0.1479</td>
</tr>
<tr>
<td>Number of Events</td>
<td>12</td>
</tr>
<tr>
<td>Observations</td>
<td>66,155</td>
</tr>
<tr>
<td>C. Ethics Exam</td>
<td></td>
</tr>
<tr>
<td>Required Passing of Ethics Exam</td>
<td>0.0063</td>
</tr>
<tr>
<td></td>
<td>(0.0052)</td>
</tr>
<tr>
<td>Mean Rate</td>
<td>0.1490</td>
</tr>
<tr>
<td>Number of Events</td>
<td>16</td>
</tr>
<tr>
<td>Observations</td>
<td>76,227</td>
</tr>
</tbody>
</table>

Notes: The table reports regressions results estimating versions of Equation 4. Panel A uses the stacked event study dataset for changes in the difficulty of the bar exam but replaces the overall simulated exclusion rate with two parts that make up the simulated exclusion rate. The “Simulated Exclusion Rate from Number of Allowed Attempts” is the exclusion rate of test takers who would have passed the bar exam had the limit on the number of allowed attempts been lifted. The “Simulated Exclusion Rate from Cut Score” is the remainder. Panel B uses a stacked event study dataset for changes in the felony ban. Panel C uses a stacked event study dataset for the adoption of the MPRE. The columns differ by the outcome: first-time discipline rate (Column 1), discipline rate (Column 2), and disbarment rate (Column 3). All outcomes are multiplied by 100. All columns control for the other policies and also include state-event, event-year, state-year, cohort, and experience fixed effects. Following Cengiz et al. (2019), standard errors are clustered at the event-by-state level. * p<0.10, ** p<0.05, *** p<0.01.
Table 4: Channels Connecting Licensing and Lawyer Quality: TWFE

<table>
<thead>
<tr>
<th>Outcome: Discipline</th>
<th>First (1)</th>
<th>Any (2)</th>
<th>Disb. (3)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Simulated Exclusion Rate from -0.0052*</td>
<td>-0.0121***</td>
<td>-0.0025</td>
<td></td>
</tr>
<tr>
<td>Number of Allowed Attempts (ppt) (0.0028)</td>
<td>(0.0043)</td>
<td>(0.0017)</td>
<td></td>
</tr>
<tr>
<td>Simulated Exclusion Rate from Cut Score (ppt) 0.0051</td>
<td>0.0217</td>
<td>-0.0143</td>
<td></td>
</tr>
<tr>
<td>Ban on Applicants with Felonies -0.0032</td>
<td>-0.0238**</td>
<td>-0.0096*</td>
<td></td>
</tr>
<tr>
<td>Required Passing of Ethics Exam 0.0037</td>
<td>-0.0107</td>
<td>0.0035</td>
<td></td>
</tr>
<tr>
<td>Mean Rate 0.1395</td>
<td>0.1994</td>
<td>0.0386</td>
<td></td>
</tr>
<tr>
<td>Observations 26,314</td>
<td>26,314</td>
<td>26,314</td>
<td></td>
</tr>
</tbody>
</table>

Notes: The table reports regressions results estimating Equation 7. The columns differ by the outcome: first-time discipline rate (Column 1), discipline rate (Column 2), and disbarment rate (Column 3). All outcomes are multiplied by 100. Standard errors are in parentheses and are clustered at the state level. * p<0.10, ** p<0.05, *** p<0.01.
Appendix A: Additional Results

Figure A1: Distribution of Discipline Rates

A. First-Time Discipline Rate

B. Discipline Rate

C. Disbarment Rate

Notes: The figure reports the distribution of the three measures of lawyer quality at the state-cohort-year level. The panels differ by the outcome used, as indicated above the panel.
Table A1: Robustness: Stack Event Study without Modifications to Cengiz et al. (2019)

<table>
<thead>
<tr>
<th></th>
<th>(1)</th>
<th>(2)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>A. First-Time Discipline Rate</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Simulated Exclusion Rate (ppt)</td>
<td>-0.0028</td>
<td>-0.0028</td>
</tr>
<tr>
<td></td>
<td>(0.0017)</td>
<td>(0.0017)</td>
</tr>
<tr>
<td>Mean First-Time Discipline Rate</td>
<td>0.1366</td>
<td>0.1366</td>
</tr>
<tr>
<td><strong>B. Discipline Rate</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Simulated Exclusion Rate (ppt)</td>
<td>-0.0067**</td>
<td>-0.0069**</td>
</tr>
<tr>
<td></td>
<td>(0.0030)</td>
<td>(0.0030)</td>
</tr>
<tr>
<td>Mean Discipline Rate</td>
<td>0.1910</td>
<td>0.1910</td>
</tr>
<tr>
<td><strong>C. Disbarment Rate</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Simulated Exclusion Rate (ppt)</td>
<td>-0.0044***</td>
<td>-0.0045***</td>
</tr>
<tr>
<td></td>
<td>(0.0013)</td>
<td>(0.0013)</td>
</tr>
<tr>
<td>Mean Disbarment Rate</td>
<td>0.0353</td>
<td>0.0353</td>
</tr>
</tbody>
</table>

**Covariates**

<p>| | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>State-Event FE</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Event-Year FE</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Other Initial Licensing Policies</td>
<td>✓</td>
<td>✓</td>
</tr>
</tbody>
</table>

Observations: 7,464 7,464

*Notes*: The table reports regressions results estimating Equation 8. The panels differ by the outcome: first-time discipline rate (Panel A), discipline rate (Panel B), and disbarment rate (Panel C). All outcomes are multiplied by 100. Following Cengiz et al. (2019), standard errors are clustered at the event-by-state level. * p<0.10, ** p<0.05, *** p<0.01.
Figure A2: Robustness: Placebo Tests

A. First-Time Discipline Rate

Notes: The figure reports results of placebo tests that drop treated states from the stacked event study dataset, randomly reassign one of the control states to be a placebo treated state in each event sub-sample, and re-estimate the primary specification. The control state that is randomly given a placebo treatment is assumed to witness the change in the simulated exclusion rate witnessed in the actual treated state for the given event. The figure reports the distribution of point estimates from 5,000 simulations for each outcome, as indicated above the panel. The red vertical line plots the actual point estimate from Column 3 of Table 2.
Table A2: Robustness: TWFE

<table>
<thead>
<tr>
<th></th>
<th>(1)</th>
<th>(2)</th>
<th>(3)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>A. First-Time Discipline Rate</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Simulated Exclusion Rate (ppt)</td>
<td>-0.0048</td>
<td>-0.0052*</td>
<td>-0.0051*</td>
</tr>
<tr>
<td></td>
<td>(0.0029)</td>
<td>(0.0029)</td>
<td>(0.0028)</td>
</tr>
<tr>
<td>Mean First-Time Discipline Rate</td>
<td>0.1395</td>
<td>0.1395</td>
<td>0.1395</td>
</tr>
<tr>
<td><strong>B. Discipline Rate</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Simulated Exclusion Rate (ppt)</td>
<td>-0.0102**</td>
<td>-0.0108**</td>
<td>-0.0118**</td>
</tr>
<tr>
<td></td>
<td>(0.0048)</td>
<td>(0.0049)</td>
<td>(0.0044)</td>
</tr>
<tr>
<td>Mean Discipline Rate</td>
<td>0.1994</td>
<td>0.1994</td>
<td>0.1994</td>
</tr>
<tr>
<td><strong>C. Disbarment Rate</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Simulated Exclusion Rate (ppt)</td>
<td>-0.0023</td>
<td>-0.0025</td>
<td>-0.0026</td>
</tr>
<tr>
<td></td>
<td>(0.0017)</td>
<td>(0.0017)</td>
<td>(0.0016)</td>
</tr>
<tr>
<td>Mean Disbarment Rate</td>
<td>0.0386</td>
<td>0.0386</td>
<td>0.0386</td>
</tr>
<tr>
<td><strong>Covariates</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>State-Year FE</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Cohort FE</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Experience FE</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Other Initial Licensing Policies</td>
<td>✓</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Observations</td>
<td>26,314</td>
<td>26,314</td>
<td>26,314</td>
</tr>
</tbody>
</table>

*Notes: The table reports regressions results estimating Equation 7. The panels differ by the outcome: first-time discipline rate (Panel A), discipline rate (Panel B), and disbarment rate (Panel C). All outcomes are multiplied by 100. Standard errors are in parentheses and are clustered at the state level. * p<0.10, ** p<0.05, *** p<0.01.
Figure A3: Robustness: Leave One Event Out

A. First-Time Discipline Rate

B. Discipline Rate

C. Disbarment Rate

Notes: The figure reports estimates on the simulated exclusion rate from Column 3 of Table 2 while leaving out one state at a time. Coefficients are reported as circles and 90 percent confidence intervals are reported as lines. The results in blue are the leave-one-event-out estimates, and the results in thick red are the primary results without leaving out any event. The panels differ by the outcome used, as indicated above the panel.
Table A3: Measuring Bar Exam Difficulty by Cut Score and Number of Allowed Attempts

<table>
<thead>
<tr>
<th>Outcome: Discipline Rate</th>
<th>First-Time</th>
<th>Any</th>
<th>Disbarment</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(1)</td>
<td>(2)</td>
<td>(3)</td>
</tr>
<tr>
<td><strong>A. Linear Cut Score and Limited Attempts</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cut Score</td>
<td>-0.0005</td>
<td>-0.0001</td>
<td>-0.0004</td>
</tr>
<tr>
<td></td>
<td>(0.0010)</td>
<td>(0.0012)</td>
<td>(0.0013)</td>
</tr>
<tr>
<td>Attempts Limited</td>
<td>-0.0050</td>
<td>0.1941</td>
<td>-0.0138</td>
</tr>
<tr>
<td></td>
<td>(0.0084)</td>
<td>(0.1513)</td>
<td>(0.0137)</td>
</tr>
<tr>
<td>Cut Score</td>
<td>-0.0015</td>
<td></td>
<td>-0.0034**</td>
</tr>
<tr>
<td>× Attempts Limited</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>B. Top Quartile of Cut Score and Limited Attempts</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Top Quartile Cut Score</td>
<td>-0.0037</td>
<td>-0.0039</td>
<td>-0.0135</td>
</tr>
<tr>
<td></td>
<td>(0.0084)</td>
<td>(0.0088)</td>
<td>(0.0121)</td>
</tr>
<tr>
<td>Limited Attempts</td>
<td>-0.0050</td>
<td>-0.0051</td>
<td>-0.0141</td>
</tr>
<tr>
<td></td>
<td>(0.0084)</td>
<td>(0.0105)</td>
<td>(0.0137)</td>
</tr>
<tr>
<td>Top Quartile Cut Score</td>
<td>0.0005</td>
<td></td>
<td>-0.0102</td>
</tr>
<tr>
<td>× Limited Attempts</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>C. Bottom Quartile of Cut Score and Unlimited Attempts</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bottom Quartile Cut Score</td>
<td>0.0007</td>
<td>0.0075</td>
<td>-0.0077</td>
</tr>
<tr>
<td></td>
<td>(0.0065)</td>
<td>(0.0076)</td>
<td>(0.0082)</td>
</tr>
<tr>
<td>Unlimited Attempts</td>
<td>0.0024</td>
<td>0.0090</td>
<td>0.0071</td>
</tr>
<tr>
<td></td>
<td>(0.0067)</td>
<td>(0.0074)</td>
<td>(0.0099)</td>
</tr>
<tr>
<td>Bottom Quartile Cut Score</td>
<td>-0.0141</td>
<td></td>
<td>-0.0263**</td>
</tr>
<tr>
<td>× Unlimited Attempts</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean Rate</td>
<td>0.1441</td>
<td>0.1441</td>
<td>0.2033</td>
</tr>
</tbody>
</table>

Notes: The table reports regressions results estimating Column 3 of Table 2 but replacing the simulated exclusion rate with various measures of the cut score and number of allowed attempts along with interactions between them. Panel A includes the cut score and “limited attempts” (an indicator for whether test takers have less than 5 allowed attempts). Panel B includes an indicator variable for the top quartile of the cut score and limited attempts as in Panel A. Panel C includes an indicator variable for the bottom quartile of the cut score and an indicator for unlimited attempts. The outcome is the first-time discipline rate in Columns 1 and 2, the discipline rate in Columns 3 and 4, and the disbarment rate in Columns 5 and 6. All outcomes are multiplied by 100. Following Cengiz et al. (2019), standard errors are clustered at the event-by-state level. * p<0.10, ** p<0.05, *** p<0.01.
Figure A4: Leave One State Out – Full Sample with Alabama

A. First-Time Discipline Rate

B. Discipline Rate

C. Disbarment Rate

Notes: The figure reports estimates on the simulated exclusion rate from estimating Equation 7 but using the full sample that includes Alabama but while leaving out one state at a time, as indicated on the y-axis. Coefficients and 90 percent confidence intervals are reported. Standard errors are clustered by state. The panels differ by the outcome used, as indicated above the panel. All outcomes are multiplied by 100.
Appendix B: Calibration and Validation Exercises for Calibrated IRT Model

Calibration. I calibrate the IRT model by making six assumptions. First, I assume that exam questions have the item-level parameters from an example published by the MBE test administrator \( a = -0.06; b = 0.56; c = 0.17 \), Albanses, 2015, Figure 1). An assumption over item-level parameters is necessary because, although the MBE test administrator has statistical standards for choosing questions during pretesting (Hill, 2015), those standards are not publicly available. Second, I follow a standard procedure in IRT calibration and assume test taker competence is normally distributed \( \theta_i \sim N(\mu, \chi) \) (Embretson and Reise, 2009, p.159). Third, I assume the error term \( \epsilon_{ij} \) is a logistic random variable with a mean of 0 and a scale parameter of 1. Fourth, I assume that the reported scaled MBE scores reflect questions answered correctly. Fifth, I assume that there is a representative distribution of test taker competency across all states. This assumption is necessary because state-level score distributions are unobserved. That said, even if the score distributions differ by state, they are endogenous to state policies, so different score distributions may still reflect the same underlying competency distribution after accounting for behavioral responses. Finally, national score distributions are not publicly available for much of my sample period (from 1980 to 2010), so I estimate the effect of changes the difficulty of the bar exam on the exclusion rate using the score distribution from the last year of my sample (2019).

Under these assumptions, \( \mu \) and \( \chi \) are the parameters to be calibrated. I calibrate these parameters to two moments based on official statistics from the 2019 bar exam (MBE National Summary Statistics, 2020): (1) the number of questions that the mean test taker answered correctly (143 questions); and (2) the standard deviation of the number of questions that test takers answered correctly (17 questions). The calibrated parameters are \( \hat{\mu} = 0.44 \) and \( \hat{\chi} = 0.58 \).

Validation. I assess the validity of the calibration in two ways. First, I use the calibrated model to simulate test taker responses on the MBE where the questions have the same properties used to calibrate the model, and then I compare the distribution of simulated total scores to the distribution of actual total scores. Because I calibrate the competence distribution rather than the total score distribution, it is possible that the calibrated model does not generate a

---

7 This assumption is required because MBE scores are scaled to produce a score that can be effectively interpreted as the number of correct answers on the 200-question MBE exam administered in 1972 (Albanses, 2015). The mean scaled score on the 2019 bar was 141. If the questions on the MBE have the same difficulty over time, this roughly translates into the average test taker getting 71 percent of answers correctly. Albanses (2015) discusses this: “For better or worse, the MBE scale we report appears to be based on the number of items answered correctly by examinees on the 200-item examination administered in July 1972 (the second-ever administration of the MBE). Even today, after 43 years, the scores from the MBE can be thought to be referenced to that early examination. It is not a direct link, however. Not one item on today’s examination or probably any examination since 1980 was on the test administered in 1972. The relationship is through the equatings that have occurred since that time. Each equating links scores to at least two earlier exams, one or more in July, the other(s) in February. Because all of these exams have been equated to earlier examinations, there is an unbroken chain of linkages back to that 1972 examination. Because the test changes its character over time and the linkages become more fragile as time passes, the base test used to serve as the reference for computing statistics employed in scaling is reset periodically. Our current base is the July 2001 examination. The difference is primarily statistical. Conceptually, the scale still harks back to the July 1972 examination.”
simulated total score distribution resembling the actual total score distribution. This is because the distribution of the total scores is not itself a normal distribution but rather the distribution that results from adding up the number of correct responses generated from Equation 1. Panel B of Figure B1 reports the total scores of the simulated and actual distributions. The figure reveals that the simulated total score distribution closely matches the actual score distribution. I test for statistical differences in the equality of the distributions using the Kolmogorov–Smirnov test and find no evidence that the distributions are different.

Second, I use the calibrated model to simulate the two year bar passage rate, and I then compare the simulated two year bar passage rate to the actual two year bar passage rate of 90 percent (Francis Ward, 2021). Two exams are offered each year, so I assume the two year bar passage rate is the percent of test takers who passed on any of four attempts. The intuition of this validation exercise is that if the model calibration is reasonable, then it should approximate not only the total score distribution on a single exam but also the eventual bar passage rates implied by it. The simulations reveal a two year bar passage rate of 90.4 percent, which closely matches the actual two year pass rate. Combined, these two validation exercises provide evidence that my model calibration approach is reasonable.
Figure B1: Calibrating Structural Model of Test Taker Exam Scores

A. Illustration of the Question-Level IRT Parameters

Notes: Panel A reports two Item Response Probability Curves to illustrate the intuition of each parameter in the IRT model of Equation 1. The discrimination parameter $a$ is the slope of the curve at the midpoint between the probability of a test taker guessing the correct answer and 1. The difficulty parameter $b$ is the test taker competency at the midpoint between the probability of a test taker guessing the correct answer and 1. The guessing parameter $c$ is the lowest probability of the curve. Panel B reports the distribution of national exam scores on the Multistate Bar Exam (MBE). The solid line labeled “Actual MBE Scores” is the kernel density function (with a bandwidth of 10) of the July 2019 MBE scaled score (MBE National Summary Statistics, 2020). The gray bars are the simulated MBE scores from the calibrated model.
Appendix C: Theoretical Model Highlighting Mechanisms

The difficulty of passing the bar exam depends on the cut score of a single exam and the number of times test takers are allowed to retake the exam. To understand the ways that these two policies can influence misconduct propensity, consider a simple model of test taker ability, bar exam scores, and misconduct. First, let competency for test taker $i$ be a function of their exogenous ability and the amount of training:

$$\theta_i(\alpha_i, c, l) = \alpha_i \sigma_{\alpha}(c, l)$$

where $c$ is the cut score, $l$ is the limit on the number of allowed attempts, $\alpha_i$ is ability, $\sigma_{\alpha}(c, l)$ is the amount of training where the sub-strict $\alpha$ assumes all test takers with the same ability respond in the same way to a set of bar exam policies. Training is a function of the cut score and number of allowed attempts because test takers may respond to more difficult exams (e.g., by studying more).

Next, let bar exam scores be a function of underlying competency and noise:

$$s_i(c, l) = \theta_i(\alpha_i, c, l) + \epsilon_i$$

Test takers with scores $s_i(c, l) > c$ pass the exam. Because of the randomness in test takers’ scores and because test takers could improve their score after failing it, some test takers pass only after failing it. Given this setup, let $y_{\alpha}(c, l)$ be the rate that test takers with ability $\alpha$ pass the exam on any attempt.

Next, assume misconduct propensity is a function of competency, whether a test taker failed the bar exam at least once before passing it, and random noise:

$$m_i(\alpha_i, c, l) = \theta_i(\alpha_i, c, l) + 1_i(c, l) + \kappa_i$$

where $1_i(c, l)$ is an indicator function for whether the test taker failed the bar exam before passing it and $\kappa_i$ is random noise.

Finally, test takers can respond to a more difficult exam by opting out of taking the exam, such as by taking the bar exam in another state. This means that bar exam difficulty can change the distribution of the ability of test takers in the state. Let the density of test takers with different competencies in the population be endogenous to the difficulty given by $p_{\alpha}(c, l)$.

Under this setup, the population misconduct rate is given by

$$\bar{m}(c, l) = \int_{\alpha = -\infty}^{\alpha = \infty} \int_{s = c}^{\infty} y_{\alpha}(c, l)p_{\alpha}(c, l)m(\alpha, c, l)$$

In this setup, there are three ways that the difficulty of the bar exam can affect misconduct. First, exam difficulty can affect misconduct by excluding test takers with different competencies ($y_{\alpha}(c, l)$). Second, exam difficulty can affect misconduct by affecting the distribution of test takers with different competencies in the population ($p_{\alpha}(c, l)$). Third, exam difficulty can affect test takers’ underlying rate of misconduct ($m(\alpha, c, l)$), including by affecting the share of test takers who fail before passing and by affecting how much test takers prepare for the exam.
Appendix D: Calibration for Decomposition

To estimate quitting behavior, I use descriptive statistics on the number of attempts and passers at each attempt from Wightman (1998). I then calibrate the behavioral response parameter on improvements in scores after subsequent attempts ($\lambda$) by assuming that it has the functional form in Equation 5 and an exogenous quit rate. Under these assumptions, I search for the parameter $\lambda$ that minimizes the squared distance between the share of lawyers who pass on different attempts (conditional on failing at least once).

Figure D1 reports descriptive statistics of the additional parameters for delayed entry from failing the bar exam. Panel A reports the quit rate after each subsequent attempt. For test takers who fail the bar exam the first time, 3 percent do not retake the exam. For test takers who fail a second and third time, 14 percent and 30 percent do not retake the exam after the attempt, respectively. Panel B reports the actual and simulated shares of lawyers who pass on repeated attempts for different values of $\lambda$. As the figure reveals, the assumed functional form matches the data well. The calibrated parameter is $\lambda = 0.61$. 
Figure D1: Estimating Parameters for Quitting Behavior and Improvement in Scores After Failing the Bar Exam

A. Quitting Behavior

B. Improvement of Bar Exam Score

Notes: Panel A reports the breakdown of test takers, failed attempts, and the implied quit rate after each subsequent attempt ($\alpha$) using data from Wightman (1998). Panel B reports the calibration of the parameter for how much test takers who fail the bar exam improve their scores on subsequent attempts ($\lambda$).