1-3-2023

JD-Next: A Valid and Reliable Tool to Predict Diverse Students' Success in Law School

Jessica Findley

Adriana Cimetta

Heidi Burross

Katherine Cheng

Matt Charles

See next page for additional authors

Authors

Jessica Findley, Adriana Cimetta, Heidi Burross, Katherine Cheng, Matt Charles, Cayley Balser, Ran Li, and Christopher Robertson

DOI: 10.1111/jels.12342

ORIGINAL ARTICLE

JOURNAL OF EMPIRICAL LEGAL STUDIES

JD-Next: A valid and reliable tool to predict diverse students' success in law school

Jessica Findley[1] | **Adriana Cimetta**[1] | **Heidi Legg Burross**[1] | **Katherine C. Cheng**[1] | **Matt Charles**[1] | **Cayley Balser**[1] | **Ran Li**[1] | **Christopher Robertson**[1,2]

[1]University of Arizona, Tucson, Arizona, USA

[2]School of Law, Boston University, Boston, Massachusetts, USA

Correspondence
Christopher Robertson, Department of Law, Boston University, 765 Commonwealth Ave, Boston, MA 02215, USA.
Email: ctr00@bu.edu

Abstract

Admissions tests have increasingly come under attack by those seeking to broaden access and reduce disparities in higher education. Meanwhile, in other sectors there is a movement towards "work-sample" or "proximal" testing. Especially for underrepresented students, the goal is to measure not just the accumulated knowledge and skills that they would bring *to* a new academic program, but also their ability to grow and learn *through* the program. The JD-Next is a fully online, noncredit, 7- to 10-week course to train potential JD students in case reading and analysis skills, prior to their first year of law school. This study tests the validity and reliability of the JD-Next exam as a potential admissions tool for juris doctor programs of education. (In a companion article, we report on the efficacy of the course for preparing students for law school.) In 2019, we recruited a national sample of potential JD students, enriched for racial/ethnic diversity, along with a sample of volunteers at one university ($N = 62$). In 2020, we partnered with 17 law schools around the country to recruit a cohort of their incoming law students ($N = 238$). At the end of the course, students were incentivized to take and perform well on an exam that we graded with a standardized methodology. We collected first-semester grades as an outcome variable, and compared JD-Next exam properties to legacy exams now used by law schools (the Law School Admissions Test (LSAT), including converted GRE scores). We found that the JD-Next exam was a valid and reliable predictor of law school performance, comparable to legacy exams. For schools ranked outside the Top 50, we found that the legacy exams lacked significant incremental validity in our sample, but the JD-Next exam provided a significant advantage. We also replicated known, substantial racial and ethnic disparities on the legacy exam scores, but estimate smaller, non-significant score disparities on the JD-Next exam. Together this research suggests that, as

an admissions tool, the JD-Next exam may reduce the risk that capable students will be excluded from legal education and the legal profession.

INTRODUCTION

Both law schools and prospective students need tools to predict the likelihood of the students' success in law school. Undergraduate grades and standardized tests, namely the Law School Admissions Test (LSAT) and the GRE, are the primary tools used today. According to one recent study with 21 law schools and 1587 students, when undergraduate grade point average (UGPA) is combined with either the GRE or LSAT scores, they predicted 42% of the variation in law school grades (Klieger et al., 2016, at 9). Based on this incomplete information, admissions officers are left without a quantitative measure of the remaining variance in law school grades. Accordingly, some accepted students will do much worse than predicted, while other rejected students could have done much better than predicted.

Undergraduate grades are an imperfect predictor of law school performance for at least two reasons. First, undergraduate studies are different than professional studies for the JD. Although general intelligence and earnestness may apply to both undergraduate and professional legal studies, more specific learned skills may not translate across the fields. For example, successfully determining the entropy change in a chemical reaction may be necessary to excel on a chemistry exam, but that same ability may not be useful for discerning the rule of law from an 1800s property case as a first-year law student. Second, undergraduate courses of study and the grading practices in those courses of study are heterogenous from each other—across majors within universities, between universities, and between countries for international students. Using a single metric of UGPA to make comparisons across students with heterogenous programs of study and grading systems is very difficult. In particular, students in science majors are thought to be disadvantaged by stricter departmental grading practices (Goldman et al., 1974, p. 356). In a 2011 study at one university, the authors concluded that "the grading policies used by [students] instructors were nearly as important in determining their GPA and class rank as was their academic performance" (Johnson, 2003). On the other hand, a 2016 study of 1400 student records from 2 law schools found that college major is a significant predictor of law student grades, with both STEM (science, technology, engineering, math) and EAF (economics, accounting, finance) majors being advantageous (Marks & Moss, 2016). More sophisticated modeling of UGPA may be able to solve for some of this heterogeneity.

Standardized tests date back to the Han dynasty in China (Zwick, 2013). For university admissions, some histories trace back to 18th-century France, where the idea of admitting students "based on test scores rather than privilege, was certainly compatible with the principles of equality that characterized the French Enlightenment" (Zwick, 2013, p.11). In the United States, it is often said

JOURNAL OF EMPIRICAL LEGAL STUDIES

that standardized tests were introduced to reduce unfair advantages given to the mostly white, male, wealthy, Protestant students attending more elite schools (Shultz & Zedeck, 2012, pp. 51, 53). However, in a noted 2015 book, Lani Guinier criticized "The testocracy [as] a twenty-first-century cult of standardized, quantifiable merit [that] values perfect scores but ignores character." (Guinier, 2015). Guinier argued that standardized tests tend to measure race and socio-economic status and argued for a reorientation that values "democratic merit."

Today, some compellingly argue that "overreliance" on standardized tests tends to disadvantage the populations that are underrepresented in law schools (Kidder, 2000; Sampson & Boyer, 2001; Taylor, 2019; Taylor, Scott, & Jackson, 2021). Indeed, the Law School Admissions Council (LSAC), which provides the LSAT, reported in 2014 that "African American" (Black) test takers scored a 142 on average, while that "Caucasian" (White) test takers scored a 153 on average (an 11-point difference, which was nearly double the standard deviation) (Dalessandro et al., 2014; see also Dalessandro et al., 2000). Disparities were also shown for other groups. In evaluating this debate, one must distinguish lower test scores that accurately predict risk of educational failure for some students (likely due to larger systemic factors in the society), versus lower test scores that may be due to biases in the test itself, if it were to fail to accurately predict the future of some students versus others (Burton & Wang, 2005; Wightman, 2000).

The American Bar Association (ABA) accredits law schools. Unlike other accreditors of graduate and professional schools, the ABA mandates that every "law school shall require each applicant for admission as a first-year JD degree student to take a valid and reliable admission test to assist the school and the applicant in assessing the applicant's capability of satisfactorily completing the school's program of legal education" (ABA, 2016). Nonetheless, the chair (at the time) of the ABA Council's Standards Review Committee called the mandate "rather toothless" because it does not include any minimum test score (ABA, 2017). The ABA's own official Interpretation 503-2 confirms that the Standard "does not prescribe the particular weight that a law school should give" to a test score when deciding whether to admit a student (ABA, 2016). Nonetheless, the ABA Council has sent letters threatening the accreditation of schools who have attempted to use other admissions pathways, which do not rely on one of the approved standardized tests or an approved "variance" from that standard (Binno v. ABA, 2016). In litigation challenging the testing mandate for having a discriminatory effect under the American Disabilities Act, courts have held that even if those allegations were true, law school aspirants lack standing to sue the ABA (Binno v. ABA, 2016).

While the ABA's testing mandate targets "satisfactory completion" of the JD program, this is not a substantial problem for most students at most law schools, and is not the metric typically used to validate admissions tests (Robertson et al., 2022). The ABA exam mandate serves another function, however. Law schools carefully manage their own LSAT medians to preserve their US News law school rankings, which effectively sorts students into the hierarchy of ranked law schools (Curtis, 2019; Haddon & Post, 2006; Kidder, 2000).

Lower-scoring students will generally pay more at the same schools or attend lower-ranked schools (Organ, 2017; Whitford, 2017; Zirkel, 2019).

One might worry about disparate levels of access to and use of test preparation programs, which can be expensive in terms of time and money (Amabebe, 2020). Major test providers, such as LSAC and ETS, now also offer some free test prep programs services, while also arguing that test preparation has limited effects. According to one review, "they typical magnitude for coached preparation is about 25% of one standard deviation" on standardized test scores (Kuncel & Hezlett, 2007, p. 1081).

In 2018, several law school deans organized to repeal the requirement of standardized tests (Miller et al., 2017). The ABA Council voted to revoke the requirement (ABA, 2018). However, the ABA House of Delegates, which sets final policy for the accreditor, did not look favorably on the proposal, and it was withdrawn. As of 2022, the ABA Council as again called for comments on changes to the testing mandate (Martinez et al., 2022).

As quoted above, the ABA Standard requires that the tests used by accredited schools be valid and reliable. The ABA Standards and their official interpretations provide neither specific criteria nor thresholds for determining validity or reliability, nor do they provide a process or mechanism for approving new tests (Martinez et al., 2022; ABA, 2016, p. 33). However, the official interpretations characterize the LSAT as something of a default, requiring that schools using other tests "shall demonstrate that [any] other test is a valid and reliable test to assist the school in assessing an applicant's capability to satisfactorily complete the school's program of legal education" (ABA, 2016, p. 33; see also Thomas, 2003). In 2020, LSAC began offering a new test, the LSAT-Flex, but there has been no public disclosure of validity data or public action by the Council to accept that test. A number of older studies have examined the predictive validity of the original LSAT in relation to law school GPA (LGPA; Shultz & Zedeck, 2011).

In 2016, University of Arizona (UArizona) partnered with ETS to test the validity and reliability of the GRE General Test for law school admissions (Klieger et al., 2016; Randazzo, 2016). They found statistically and practically significant levels of validity and reliability, and UArizona began using the test for admissions (Klieger et al., 2016, p. 14; Randazzo, 2016) In December 2021, relying on a national validity study with 21 law schools, the ABA Council affirmed the acceptability of this use (Jaschik, 2021). The GRE is now accepted by approximately 100 law schools, allowing students to choose which test to take (Educational Testing Services, 2021).

The LSAT and GRE General exams include reading comprehension, analytical reasoning, and logical reasoning questions, though neither test focuses specifically on legal materials. The proven validity of these exams suggests a correlation between these general skills and specific skills needed in law schools, at least to the extent that those skills are measured in first-year grades. Yet, these exams have two potential limitations. First, they do not directly measure the specific applied skills that students will need to exercise on day one of their

JOURNAL OF EMPIRICAL LEGAL STUDIES

JD program—reading classic cases, extracting legal doctrine, and applying it to new facts as presented on an exam. Second, these tests do not directly measure the ability of students to *learn* those skills when placed in a well-defined environment similar to law school, with appropriate scaffolding to learn and formative feedback along the way.

The first limitation of these traditional forms of testing is in their gap between the performance measured on the test and the performance measured in law school. In short, these tests look nothing like law school exams, so it is remarkable if they have predictive power for that purpose. Instead, these tests may be measuring a third thing (or set of things), which happens to be more or less correlated with both test performance and law school performance.

Reflecting on this insight, which seeks to close the measurement-performance gap, athletic coaches and employers rely on tryouts, talent auditions, or "work sample tests," where a potential employee is given a task (e.g., writing a computer code to solve a given problem) and asked to perform it (Tucker, 2016; Den Hartigh et al., 2018). In 1998, reiterating seminal work, Schmidt and Hunter reviewed 19 different tools for predicting job performance, and found that when compared to a measure of general mental ability, work sample tests had greater predictive power (Schmidt & Hunter, 1998, p. 262; see also Roth et al., 2005) This "work-sample" approach suggests that if law schools want to predict how well a student is able to learn new legal material and apply law to facts—then they should observe applicants doing exactly that.

In a systematic review published in *Science*, Kuncel and Hezlett point out that most standardized tests assess a combination of language ability, quantitative ability, writing ability, and analytical reasoning ability or specialized knowledge (Kuncel & Hezlett, 2007, p. 1080). "Although the general verbal and quantitative scales are effective predictors of student success, the strongest predictors are tests with content specifically linked to the discipline" (Kuncel & Hezlett, 2007). Accordingly, potential students of Biology, Chemistry, Literature in English, Mathematics, Physics, and Psychology may take one of the GRE Subject tests (GRE-S) specific to those fields of study (Educational Testing Services, 2020).

A similar approach called "proximal," "trial studying," or "curriculum-sampling" testing, is becoming increasingly used in European higher education. As Niessen et al. (2018) explain,

> The rationale behind these tests is to mimic later behavior that is expected during an academic study. Thus, curriculum samples often mimic representative parts of the academic program that the student is applying to. Often, these samples are small-scale versions of an introductory course of a program, because performance in such courses is a good indicator for later academic performance (pp. 1–2).

For example, Niessen et al. (2016) evaluated such an approach to undergraduate admissions to a Psychology program, and found that the proximal test was consistently the best predictor of students' academic performance during their first year. Because this kind of test is closer to real learning challenges and situations the future student will encounter, it can be a good predictor of academic performance (Farley et al., 2019). In addition, the lived experience preparing for and taking the test allows students to develop self-knowledge and self-select into the program if it seems like a good match. Consistent with these principles of proximal testing, we designed JD-Next as a bridge program to be similar to actual law school, and we designed it in concert with the development of a standardized test. In this way, JD-Next is like a tryout for law school.

The second, and arguably more important, limitation of traditional admissions tests is that they measure students' intellectual abilities, a form of human capital developed at a particular point at the time of testing, given the educational experiences and other capital that potential students' households have accumulated up to that point (Sternberg & Grigorenko, 2020, pp. 21–22). Psychologists have been aware of this limitation of traditional "static" tests since the earliest days of the field, but the problem was crystalized in the 1930s by Russian psychologist Lev Vygotsky, who introduced the concept of "dynamic testing" to modern psychology (Lidz, 1995). In this paradigm, teaching and feedback are considered learning interventions, which are essential to the learning process and thus of the measurement of learning potential. As Sternberg and Grigorenko (2020) explain,

> [D]ynamic testing is based on the link between testing and intervention and examines the process of learning as well as its products. … In other words, what is tested in not just previously acquired skills, but the capacity to master, apply, and reapply skills taught in the dynamic testing situation. This view of the testing procedure underlies the use of the term, *test of learning potential*, which is often applied to dynamic testing (p. 29).

The key question for an admissions dean choosing an admissions test is whether to measure the skills and knowledge that *has been* learned, or the skills and knowledge that *could be* learned with the benefit of their educational program.

A dynamic approach to testing has important social, normative, and even political aspects for diversity, inclusion, and fairness. If traditional asset-testing is conducted in a society that suffers from a disparity of background capital and opportunities, then the resulting maldistribution of human capital will be measured by static tests (Conley, 2009). In such an unequal society, it would be unsurprising to find disparities in static tests that correlate with underlying economic inequalities, which are correlated with racial, ethnic, and geographic lines (Reardon et al., 2019). Admissions practices that selected students at least partly on the basis of such tests would reproduce such hierarchies across generations. Instead, as Lani Guinier has been quoted as arguing, “We can alter how we think about merit, from something a child is born with to something that she (and/or we) can help cultivate” (Jaschik, 2015). This perspective echoes the original motivation for dynamic testing, which “was viewed as opening up the world for the child, whereas static testing was viewed as closing it” (Sternberg & Grigorenko, 2020, p. 37).

If education, and law school in particular, has transformative potential, then it should be wary of selecting for those who have *already* developed the key skills that it seeks to inculcate. Higher education should, instead, select for potential, since the school is in the very business of actualizing that potential, not merely placing laurels on those who are already more successful.

The notion of dynamic, proximal testing is no panacea. How to operationalize it for law schools is a key practical question. And whether such testing of learning potential would have additional predictive value for law school performance is an empirical question.

METHODS

We sought to test whether the JD-Next exam is a valid and reliable predictor of law school performance, and whether it provides incremental predictive value at schools with varying levels of selectivity. In addition, we sought to determine whether the test suffers from score disparities which would disadvantage underrepresented groups of students, if the test were used for admissions decisions. In this part, we provide an overview of the JD-Next course curriculum, the construction and grading of the exam, our research design including incentives, and the populations of potential JD students that were recruited to participate.

The course and exam development

The JD-Next course is described in detail in the companion article (Cheng et al., 2021). In short, in 2019 we started with a scaffolded fully online pedagogy, consisting of 15 doctrinal classes covering 18 Contracts law cases and 8 skills workshops across 7.5 weeks. Each week consisted of two classes and one skills workshop. The doctrinal law classes drew on the kinds of cases that a law student could expect to encounter in a 1L (first year law student) Contract law course, and skills workshops each introduced a skill, for example, how to identify the rule in a case, with a short 3- to 8-min video explaining the skill and an example of the skill being exercised. The course was designed around the idea that, while students needed some doctrinal material to work with, the development of skills was the key goal (Christensen, 2007; Christensen, 2009). The course was asynchronous, so students could complete assignments at times that worked best for their schedules. Nonetheless, we offered a recommended pace of completion to help students stay on track.

In 2019, we offered several types of incentives to participants in the program. The primary incentive we offered was $25 to complete the course and $75 for submitting their first semester grades. To address attrition issues, we also offered students a $45 bonus using a "banking" scheme, where students banked bonuses between $5 and $15 for each week completed. Once students completed the course, they received the cash bonus. Throughout the course, we offered additional incentives to encourage students to stay on track, which included textbooks (2), law school t-shirts (4), and an iPad (1). Students who were on track at the time of the drawing receive an entry into the drawing. To take and perform on the exam, students were entered into a drawing for $125 prizes a number of times based on the students' rank in the exam performance in their respective course section (i.e., for each exam question correctly answered, the student would have an additional chance to win the drawing). Similar incentives were offered for the 2020 cohort.

The JD-Next exam included multiple choice questions and an essay question. A strategic development process ensured balanced coverage of the course material, including both the skills and the doctrinal material. Initially, we created three questions over each skill and each of the 15 cases ("topics") covered in the doctrinal content. After creating 60 questions, we reviewed them qualitatively as a research team, with the assistance of subject-matter experts (in Contracts) at the University of Arizona College of Law. In this review process, we excluded 16 questions based on various concerns and, we were left with 44 acceptable questions. Questions were eliminated if they were overly simple, overly complicated or nuanced, incorrect representation of the law, or duplicative of the material that was covered by other questions. In the final version of the exam, most topics contained two or three questions.

The essay question was drafted by Professor Robert Williams, the creator of the original undergraduate course, and covered 5 topics, worth 15 points overall. Accordingly, the 2019 test was scored out of 59 points total (44 multiple choice plus 15 essay).

We created a grading rubric for the essay to support interrater reliability. After creating the initial rubric, the graders completed six cycles of review (approximately 30 h of training). Each cycle included grading 10–20 essay samples, writing detailed notes, discussing differences in decision-making, and further refining the rubric to ensure consistency across graders. The ultimate grading did not begin until the graders achieved greater than 85% interrater reliability.

For the 2020 cohort, we made various revisions to the course based on qualitative feedback and quantitative analyses, described in Section 3.2 below. To reduce the intensity of the experience for students over their summers, the

JOURNAL OF EMPIRICAL LEGAL STUDIES

JD-Next course ran over 10 weeks and was reduced to eight Skills Workshops and nine doctrinal classes. Each week consisted of two classes except for Week 10, which had one class and the final exam. The course was scaffolded to first focus on skills development of case briefing and legal analysis during the first 4 weeks. The schedule offered a break in Week 5, which students could use as a time to catch up on work on the first half of the course before beginning the applying the new skills to doctrinal material in Week 6. The remaining 5 weeks were spent on 9 doctrinal law classes covering 11 classic cases found in 1L Contracts.

The 2020 participants were also incentivized to take the JD-Next exam, as in 2019. The exam was expanded to 60 multiple choice questions and, again, one essay question, we revised or replaced questions that performed poorly in 2019 according to their difficulty (percent of respondents answering correctly) and discrimination (point-biserial scores). Scoring of the exam again consisted of multiple-choice questions receiving 1 point for each correct answer and the essay worth 15 points for a total score of 75 points. We implemented a similar process of training essay graders and retraining until they reached 85% interrater reliability.

For the 2019 cohort, the average score was 41.08 ($SD = 6.31$), which is 69.6% of the 59 possible points scored. For the 2020 cohort, the average score was 51.98 ($SD = 11.15$), which is 69.3% of the 75 possible points scored. The two annual averages are almost identical, notwithstanding the replacement of questions and the addition of new questions (along with a different sampling strategy described below).

Research design and population

In the first effort to field the JD-Next program in 2019, we used a three-group, partially randomized, experiment. A National sample was recruited into a blinded study, and then randomized to a treatment and an active control (placebo group), though our primary analyses here focus on the treatment group, exposed to the JD-Next course described above.[1] In addition, a sample from one university (UArizona) was recruited without blinding, and they self-selected into the treatment. The research protocol was determined to be exempt by the Institutional Review Board at University of Arizona, and all participants provided informed consent.

Participants for the National cohort were identified using the Law School Admissions Council prospective student database and recruited by direct email.

[1]Students in the placebo participated in a 7.5 week course, writing weekly essays over law-related TV shows. The placebo control is primarily relevant to our companion paper, testing for course efficacy in improving law school performance (*see* Cheng et al, 2022), but is mentioned below, in reference to content validity of the exam. Admission to the course was stratified by race, such that as participants within each ethnic group were admitted to the treatment and control conditions.

Our inclusion criteria required either being admitted to a law school in the coming fall semester, or for under-represented students, being waitlisted at a law school. Given the goal to understand whether JD-Next had strong efficacy and predictive power for under-represented minorities in particular, of the 11,587 invitations we sent, we oversampled Native American or Native Hawaiian (137 invitations sent, 1% of the total invitations) Asian (1639, 14%), Black or African American (2798, 24%), and Hispanic (2416, 21%) aspiring law students. To encourage the participation of Native students, we enlisted the course designer and leading Native American law professor, Rob Williams, to help write a special invitation email.

In 2019, we also recruited a sample from University of Arizona ("UArizona"). We invited all 156 students matriculating at University of Arizona to participate. Forty-five students began the course, and 25 completed the course. For purposes of these analyses, the 2019 sample includes both those recruited nationally and those recruited from UArizona.

In 2019, 69 participants completed the JD-Next exam. However, seven participants were removed from the predictive validity analysis, including four that did not have first semester LGPA and three that had JD-Next scores more than 3.5 standard deviations from the sample mean (outliers). Thus, as shown in Table 1, 62 test takers are included in the predictive validity analysis.

In its second cohort, in summer 2020, the JD-Next program invited incoming law students representing 17 schools throughout the country, who executed memoranda of understanding with JD-Next. As shown in Table 2, these schools represent a wide range of selectivity metrics, which we grouped into three roughly equal size groups. All matriculating and waitlisted first year law students (1Ls) at these 17 schools were invited to participate. After they completed their Fall semesters, each participating school provided identified grades information for each student who signed a release form.

In the 2020 cohort, we have 317 JD-Next exam scores, but 11 students did not matriculate, four did not complete the 1L semester, and 36 participants were from 2 schools that did not calculate a Fall 2020 LGPA due to the COVID-19 pandemic. Thus, missing data from 51 participants were due to systemic issues. Twenty-four participants were missing either law students' first year GPA, law school name, UGPA, or LSAT scores. Multiple attempts to obtain the missing data for the remaining 24 students from schools or directly from students were unsuccessful. Four JD-Next exam scores were more than 3.5 standard deviations below the mean and therefore not included in the analysis, using the same exclusion criterion as in 2019.

For the validity analysis, the 2020 JD-Next cohort consisted of 238 1Ls who completed the course and exam, as shown in Table 1. Eight students ultimately attended schools not originally invited to be part of the 2020 JD-Next program. Demographic information and law students' first year GPA for these students came from self-report and participant provided transcripts.

Table 1 presents the demographic data and descriptive statistics, showing substantial representation across ethnic and racial groups. As expected, due to our

 JOURNAL OF EMPIRICAL LEGAL STUDIES

TABLE 1 Descriptive statistics for 2019 and 2020 participants by annual cohort, showing % (*n*) or mean (SD)

	2019 Sample $N = 62$	**2020 Sample** $N = 238$
Gender		
Female	54.8% (34)	60.5% (144)
Male	43.5% (27)	37.4% (89)
Non-binary	1.6% (1)	1.7% (4)
Prefer not to disclose		0.4% (1)
Race		
American Indian	3.2% (2)	0.0% (0)
Asian	16.1% (10)	9.7% (23)
Black/African American	9.7% (6)	7.2% (17)
Hispanic/Latino	11.3% (7)	7.2% (17)
Multi-Race	16.1% (10)	14.4% (34)
White (non-Hisp.)	43.5% (27)	61.4% (145)
URG	38.7% (25)	25.6%% (61)
	Mean (SD)	
Age	25.55 (5.11)	25.52 (5.04)
UGPA	3.39 (0.47)	3.52 (0.40)
LSAT	158.08 (6.52)	158.62 (6.14)
LGPA	3.19 (0.57)	3.31 (0.50)
JD-Next exam	41.08 (6.31)	51.98 (11.15)

Note: URG = underrepresented groups coded here as all groups other than non-Hispanic Whites and Asians, based on statistical disparities in test-performance as shown in Section 3.4. (One test-taker in 2019 and seven in 2020 identified as White and Asian and was thereby coded as multi-race but as non-URG.) LSAT variable includes converted GRE scores. Race data are self-reported by students in our surveys (not as reported by schools from ABA). The 2019 JD-Next exam total was 59 points (44 multiple choice + 15 essay), while the 2020 exam total was 75 points (60 multiple choice + 15 essay). As percentages, the annual average scores are 69.3% and 69.6%, respectively.

Abbreviations: ABA, American Bar Association; LGPA, law school grade point average; LSAT, law school admissions test; UGPA, undergraduate grade point average; URG, underrepresented group.

sampling strategy, the 2019 cohort had greater racial and ethnic representation, with non-Hispanic Whites representing less than half of the study population (43.5%). In 2020, the research population was more similar to that of US law school matriculants (61.4% Non-Hispanic White). We also created an indicator variable "URG," to represent underrepresented ethnic and racial groups, which we here defined as all groups other than non-Hispanic Whites and Asians, based on statistical disparities in test-performance as shown in Section 3.4.

We also collected UGPA and scores for both LSAT and GRE. In 2019, we converted three students GRE scores (2.5% of the total sample) to corresponding LSAT scores, using the tool provided by ETS, thereby creating a composite variable that we call the "LSAT" for simplicity (Educational Testing Service, 2020). In 2020, eight students (3.3% of the total sample) provided GRE scores, which we also converted.

ANALYSES AND RESULTS

We report reliability and validity analyses of the JD-Next exam across 2019 and 2020 cohorts. Section 3.1 reports on our test reliability, including both the correlations across multiple choice items and the interrater reliability on essay

TABLE 2 Descriptive statistics for 2020 participants by school groupings, showing % (n) or mean (SD)

Law school clusters	Group I (Top 50)	Group II (51–100 ranks)	Group III (100+ ranks)
Number of students	93	59	86
Number of schools	8	7	9
Gender			
Female	60.2% (56)	61.0% (36)	60.5% (52)
Male	36.5% (34)	36.6% (21)	39.5% (34)
Non-binary	2.2% (2)	3.4% (2)	0% (0)
Prefer not to disclose	1.1% (1)	0% (0)	0% (0)
Race			
African American/Black	4.3% (4)	5.3% (3)	11.6% (10)
Asian	18.3% (17)	5.3% (3)	3.5% (3)
Hispanic/Latino	5.4% (5)	19.3% (11)	1.2% (1)
Multi-Race	17.2% (16)	14.0% (8)	11.6% (10)
White	54.8% (51)	56.1% (32)	72.1% (62)
Age	25.49 (4.82)	24.54(4.79)	26.19 (5.38)
JD-Next exam	54.40 (10.47)	52.19 (11.45)	49.21 (11.15)
LSAT	162.98 (4.77)	157.56 (4.07)	154.64 (5.62)
mLSAT	162.02 (1.27)	156.15 (1.36)	151.49 (2.80)
UGPA	3.59 (0.30)	3.58 (0.40)	3.41(0.46)
LGPA	3.50 (0.33)	3.16 (0.50)	3.22 (0.58)

Note: LSAT variable includes GRE score converted to LSAT. Race data are self-reported by students in our surveys. Two participants in Group II did not report their race.
Abbreviations: ABA, American Bar Association; LGPA, law school grade point average; LSAT, law school admissions test; mLSAT, median school LSAT score as reported by ABA; UGPA, undergraduate grade point average.

grading. In Section 3.2, we use correlation matrices followed by multivariate regressions to assess the predictive validity for JD-Next exam using Fall 1L grades as the key outcome. A key question is whether the JD-Next score provides a predictive advantage (incremental validity) over using other information available to admissions officers (UGPA or LSAT scores). In Section 3.3, we conduct similar multivariate regressions but disaggregate the 2020 analyses for three groupings of law schools, according to their incoming class metrics. We investigate whether the JD-Next score and/or LSAT score has incremental predictive validity across all these schools which are very different from each other. Finally, in Section 3.4, we examine whether test takers from underrepresented groups perform less well compared to White and Asian students, and whether the validity of the two exams holds up for both groups.

Reliability

Reliability can be described as "the consistency of repeated measurements of the same event by the same process" (Cronbach, 1947). Arguably, from an ex post perspective reliability is a necessary condition of validity, and thus success on the latter ensures success on the former. Nonetheless, in the development of a new exam, it is important to achieve reliable questions and grading methods, to ensure accurate measurement of underlying constructs.

Determining the reliability of the JD-Next exam was an iterative process which included assessing the internal consistency and discriminant functioning of the multiple-choice exam items and the interrater reliability among graders for the essay portion of the exam. For the 2019 cohort, all reliability analyses included exams from both the UArizona sample and the national Treatment and Control samples, total $N = 129$. (This sample is larger than in the analyses follow, because (a) outcomes data [i.e., first-year law school grades] are not necessary for reliability analyses, and (b) our validity analyses focus on the intended use of the exam for course-takers.)

First, all items of the exam were examined to assess the reliability of the total score on the multiple-choice portion of the exam by calculating a Cronbach's alpha, which was $\alpha = 0.817$. The Cronbach's alpha is an estimate of internal consistency of the scale—the law exam in this case. A Cronbach's alpha value of 0.70 with preference of 0.80 or higher is widely accepted indicator of good internal consistency in the measure (Cortina, 1993; Tavakol & Dennick, 2011).

Review of the point biserial item correlations found five items to have poor item functioning. A value of $r < 0.15$ is a commonly used threshold to identify items that poorly discriminate between high and low performers on the exam (Ebel & Frisbie, 1991).

The internal consistency reliability analysis was rerun with these five items excluded resulting in a slightly higher Cronbach's alpha, $\alpha = 0.852$. These items were revised for the 2020 exam.

The internal consistency of the two constructs (standard Contract law content knowledge and foundational skills in case-reading and analysis knowledge) underlying the exam was also examined. Based on the previous discriminate item functioning analysis, the five problematic items were removed from the analysis. The subscale for standard Contracts law content contained 26 items with a good Cronbach's alpha of 0.805. The subscale for the foundational skills in case-reading and analysis contained 13 multiple-choice items with a Cronbach's alpha of 0.669, which is slightly lower than the acceptable level for good internal consistency. The reliability of potential subscales will be reexamined after item revisions are made and data are available in the second implementation of the JD-Next program.

The essay question was manually scored by two trained graders. To ensure the scoring was consistent across graders, a grading rubric was created be the course instructor, law professors, and educational psychologists. Graders used the rubric to each grade 30 essay questions (23%) of exams across two rounds of training. This training process was iterative and the operational definitions in the rubric were revised as questions in grading surfaced. At the completion of the training, the two graders achieved an inter-rater reliability of 95.9% when using the finalized rubric to score the essay question. After training and calibration was complete, the training exams were then reinserted into the stack and regraded for subsequent analyses.

Based on the findings of the reliability analyses conducted on the JD-Next exam, results indicate the exam scores are reliable as evidenced by acceptable internal consistency and very high interrater reliability in scoring the essay. We undertook similar efforts when revising the exam for 2020.

Overall validity of the JD-Next exam

In this section, we describe our findings for the psychometrics of the exam, focusing on several aspects of validity. The primary questions are whether the exam represents the ideas that it is supposed to test and whether it predicts performance in law school.

First, we considered content validity, to ensure that the exam was representative of the course as it was taught (Kubiszyn & Borich, 2003). After the exam items were written, law professors and instructional staff with experience and expertise in Contracts law reviewed the items to determine if the exam was representative of the standard contract and foundational skills constructs the exam was designed to measure. Through this verification of exam content by experts, the exam was found to cover the type of content found in a typical first year

Contracts law course offered at most US law schools, though in narrowed scope. Thus, content validity of the exam was established.

Second, we considered construct validity, to ensure the exam measured what was intended. The exam was developed by a group of stakeholders with expertise in Contracts law (Contracts law professors), foundational skills in case-reading and analysis (law professors and instructional staff), and assessment (educational psychologists). The main constructs covered in the JD-Next Prep course represent the content and skills typically covered in first-year, first semester law school. The exam questions were deliberately written to assess either contract content or the skills and analysis required in reading cases.

Specifically, the initial 45-item test was developed with 44 multiple-choice items, 30 items primarily assessing standard contract content knowledge and 14 items primarily assessing foundational skills in case-reading and analysis knowledge, and 1 essay question focused on the application of case-reading and analysis skills used for legal writing. Experts examined the underlying traits/constructs (content and skills) to ensure that the skills being measured by the JD-Next exam were interpreted accurately.

To understand the relationships between predictor and outcome variables, with concerns for multicollinearity, we first examined the correlations for all variables used in the regression models, shown in Table 3. There are moderate relationships among the many of the variables. Not surprisingly, the LSAT score and median LSAT score for a school were strongly correlated, reflecting the use of those tests for sorting students into ranked law schools. For each regression shown below, the variance inflation factor (VIF) value was used to determine if multicollinearity was a problem. If the VIF was >2.5, then multicollinearity was reported as a problem in the regression model.

Third, we considered predictive criterion-related validity by using the JD-Next exam scores to predict first semester LGPA. A correlation of 0.50–0.70 serves as the threshold indicating strong predictive validity (Kubiszyn & Borich, 2003). We focus on the results for the combined treatment groups from 2019 ($N = 62$), reflecting that the JD-Next exam is designed to measure ability to learn the material taught in the associated course.[2] The unadjusted correlation ($r = 0.48$) is shown in Figure 1a.

We seek to understand not just whether the exam provides a valid prediction, but also whether it provides incremental value above and beyond the other

[2]In separate analyses with the same regression models, not shown, we examined the predictive validity of the exam for students in the UArizona subsample ($N = 24$), and found strong predictive power ($r = 0.705$) and a significant improvement over using UGPA alone ($p < 0.001$). For the 2019 control group, who took the JD-Next exam after participating in the placebo course, which consisted of writing short essays about law-related television shows ($N = 57$), we found strong predictive power ($r = 0.52$), but the base model of UGPA and median LSAT explained a surprisingly large amount of the variance ($r = 0.33$) and the marginal contribution of adding the JD-Next score was non-significant ($p = 0.068$). For the 2019 full sample of test takers ($n = 119$), regardless of whether they took the associated JD-Next course, we found strong predictive power ($r = 0.532$) and a significant improvement over using UGPA alone ($p < 0.001$).

TABLE 3 2019 Cohort correlation matrix of variables included in the user-determined stepwise regression analysis of students' first semester LGPA ($N = 62$)

	Median LSAT	LSAT	JD-next exam	LGPA
UGPA	0.345**	0.095	0.202	0.084
Median LSAT		0.607**	0.266*	0.354**
LSAT			0.528**	0.689**
JD-Next exam				0.480**

Note: LSAT variable includes converted GRE scores. This sample includes both the Treatment and Control groups from the 2019 experiment, along with the UArizona Treatment group.
Abbreviations: ABA, American Bar Association; LGPA, law school grade point average; LSAT, law school admissions test; UGPA, undergraduate grade point average; UArizona, University of Arizona.
*$p < 0.05$; **$p < 0.01$.

information that would be available to a law school admissions office, which would include UGPA and the median LSAT score, as a measure of the general selectivity and competitiveness of the class (Hunsley & Meyer, 2003, pp. 446–455).

Taking account of other variables, we used a series of linear and stepwise regression analyses with a predetermined order of variable entry, as shown in Table 4. The median LSAT score for the participants' school was included in the regression models to account for the differing selectivity and rigor of the participants' law schools, and UGPA was used as a primary piece of information available to admissions officers. Model 1 included the JD-Next exam score as the independent variable predicting LGPA, the dependent variable. Some readers will be interested in how this new exam compares to established admissions procedures, and specifically whether it may be a *substitute*. It is valuable, therefore, to have an estimate of the LSAT's predictive validity within this same research sample. Accordingly, Model 2 included the traditional measure, the LSAT score, in predicting LGPA.

Alternatively, the JD-Next score may be useful as a *complement* to other standardized test scores, allowing greater predictive power even if a student already has another measure such as LSAT score (and vice versa). In Model 3, JD-Next exam score was entered followed by the LSAT score to determine whether using both scores improved prediction. The order of entry for predictor variables was reversed in Model 4 with the LSAT score entered first followed by the JD-Next exam score. Models 3 and 4 address the question of whether the JD-Next exam score can supplement or supplant the traditional law school entrance examination scores.

Model 1 shows that the addition of the JD-Next exam to the law school median LSAT score significantly predicted LGPA ($r = 0.542$) and accounted for an additional 16.7% variance in law students' first year GPA ($p < 0.001$). Model 2 shows that students' LSAT score also significantly predicted LGPA

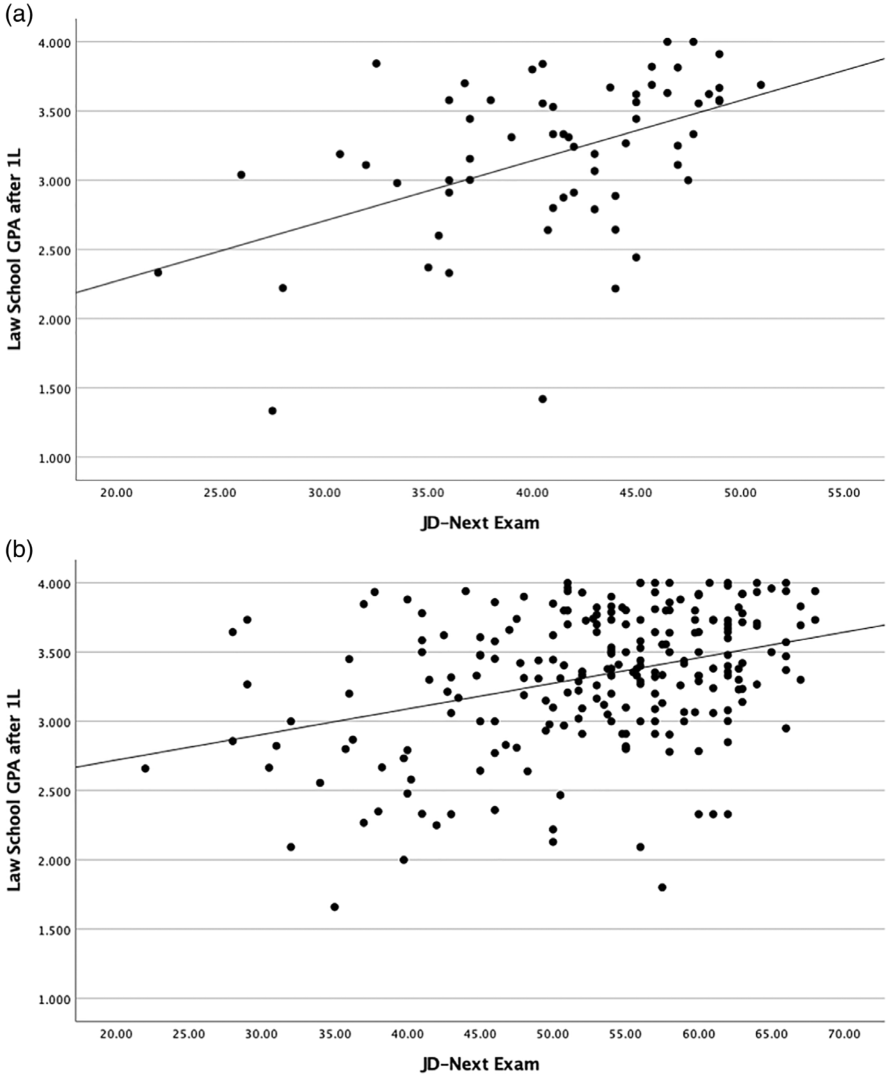

FIGURE 1 Relationship between JD-Next exam score and first semester law grade point average (LGPA). (a) 2019 Cohort, $N = 62$, $r = 0.480$. (b) 2020 Cohort, $N = 238$, $r = 0.415$

and accounted for an additional 35.8% of the variance in LGPA. When using both the JD-Next score and LSAT score to predict LGPA, order of entry into the model mattered. Model 3 shows that when entering JD-Next score into the model first followed by LSAT score, they both contributed to significantly predicting LGPA. Model 4 shows that when the JD-Next exam was entered

TABLE 4 2019 Cohort combined sample of JD-Next course takers (UArizona and National, $N = 62$), summary of user-determined stepwise regression analysis for the JD-Next or LSAT exams with UGPA and median LSAT predicting first semester LGPA

	First semester LGPA			
Variables entered	***r***	R^2	ΔR^2	ΔR^2 *p*
Model 1[a]				
UGPA and mLSAT	0.356	0.127	0.127	0.018
JD-Next	0.542	0.293	0.167	<0.001
Model 2[b]				
UGPA and mLSAT	0.356	0.127	0.127	0.018
LSAT	0.696	0.484	0.358	<0.001
Model 3[c]				
UGPA and mLSAT	0.356	0.127	0.127	0.018
JD-Next	0.542	0.293	0.167	<0.001
LSAT	0.706	0.499	0.206	<0.001
Model 4[d]				
UGPA and mLSAT	0.356	0.127	0.127	0.018
LSAT	0.696	0.484	0.358	<0.001
JD-Next	0.706	0.499	0.015	0.198

Note: LSAT variable includes converted GRE scores. Standardized regression coefficients reported in final equations.

Abbreviations: LGPA, law school grade point average; LSAT, law school admissions test; mLSAT, median LSAT for the participants' law school; UGPA, undergraduate grade point average; UArizona, University of Arizona.

[a]Final equation: LGPA = −0.096 (UGPA) + 0.273 (mLSAT) + 0.427 (JD-Next).

[b]Final equation: LGPA = 0.056(UGPA) + −0.128(mLSAT) + 0.761(LSAT).

[c]Final equation: LGPA = 0.025(UGPA) + −0.101 (mLSAT) + 0.148(JD-Next) + 0.670(LSAT).

[d]Final equation: LGPA = 0.025(UGPA) + −0.101(mLSAT) + 0.670(LSAT) + 0.148(JD-Next).

after the LSAT, it did not significantly add to the predication of LGPA for the 2019 cohort.

For the larger 2020 cohort we conducted similar analyses to determine whether JD-Next exam score predicts law students' first year GPA. We started by examining correlations, as shown in Table 5 and Figure 1b.

For the 2020 cohort, Table 6 presents our multivariate analyses for incremental predictive validity. Model 1, which includes UGPA and the median LSAT score of the schools, shows that the JD-Next exam significantly predicted LGPA ($r = 0.510$) and accounted for 26% of the variance in law students' first year GPA, a substantial incremental validity over the base model ($\Delta R^2 = 0.122$, $p < 0.001$). Model 2 shows that the LSAT score also significantly predicted LGPA ($r = 0.416$) but only accounted 17.3% of the variance in law students' first year GPA, providing a smaller incremental validity

JOURNAL OF EMPIRICAL LEGAL STUDIES

TABLE 5 Correlation matrix of variables included in the user-determined stepwise regression analysis of students' first semester LGPA ($N = 238$), 2020 cohort

	Median LSAT	LSAT	JD-Next exam	LGPA
UGPA	0.262**	0.109	0.072	0.279**
Median LSAT		0.682**	0.236**	0.311**
LSAT			0.345**	0.331**
JD-Next exam				0.415**

Note: LSAT variable includes converted GRE scores.
Abbreviations: LGPA, law school grade point average; LSAT, law school admissions test; UGPA, undergraduate grade point average.
**$p < 0.001$.

($\Delta R^2 = 0.034$, $p = 0.002$). When using both the JD-Next exam and LSAT exam scores to predict law students' first year GPA, order of entry into the model mattered. Model 3 shows that when entering the JD-Next score into the model first, it significantly predicts law students' first year GPA accounting for 26% of the variance, as seen in Model 1. However, the addition of the LSAT exam score did not significantly add to the predication of law students' first year GPA ($p = 0.085$). Model 4 shows the effects of the reverse order of entry. When the JD-Next exam was entered after the LSAT exam score it significantly predicted law students' first year GPA ($r = 0.519$) and increased the variance accounted for by 9.7% ($p < 0.001$). This finding suggests that the JD-Next score may be useful as a supplemental admissions tool even for students who have taken one of the legacy exams (the LSAT ore GRE).

Our 2019 and 2020 analyses are broadly consistent in showing that the JD-Next exam is a valid predictor of law school performance, within a similar power compared to the legacy exams. The larger 2020 sample, based on the more extensive, revised exam, suggests that the JD-Next score alone provides incremental predictive power above the base model, which is highly significant, about four times the incremental value provided by the legacy exam alone. It also provides incremental, significant predictive power even for students that already have a legacy exam score.

Validity by school groupings (2020 cohort)

Among the roughly 200 law schools in the United States, there are a wide range of academic profiles. For the 2020 cohort, which recruited from a diverse group of 17 law schools, we sought to understand the validity of the JD-Next exam across that range. To allow for statistical power and to avoid identifying any particular school, we grouped the participating schools into three roughly equal sized categories, Group I: those ranked in the Top

TABLE 6 2020 Cohort ($N = 238$), summary of user-determined stepwise regression analysis for the JD-Next or LSAT exams with UGPA and median LSAT predicting first semester LGPA

	First semester LGPA			
Variables entered	**r**	R^2	ΔR^2	ΔR^2 p
Model 1[a]				
UGPA and mLSAT	0.372	0.139	0.139	< 0.001
JD-Next	0.510	0.260	0.122	<0.001
Model 2[b]				
UGPA and mLSAT	0.372	0.139	0.139	<0.001
LSAT	0.416	0.173	0.034	0.002
Model 3[c]				
UGPA and mLSAT	0.372	0.139	0.139	<0.001
JD-Next	0.510	0.260	0.121	<0.001
LSAT	0.519	0.270	0.010	0.085
Model 4[d]				
UGPA and mLSAT	0.372	0.139	0.139	<0.001
LSAT	0.416	0.173	0.034	0.002
JD-Next	0.519	0.270	0.097	<0.001

Note: LSAT variable includes converted GRE scores. Standardized regression coefficients reported in final equations.

Abbreviations: LGPA, law school grade point average; LSAT, law school admissions test; UGPA, undergraduate grade point average.

[a]Final equation: LGPA = 0.209 (UGPA) + 0.171 (mLSAT) + 0.359 (JD-Next).

[b]Final equation: LGPA = 0.232 (UGPA) + 0.077 (mLSAT) + 0.254 (LSAT).

[c]Final equation: LGPA = 0.219 (UGPA) + 0.081 (mLSAT) + 0.332 (JD-Next) + 0.138(LSAT).

[d]Final equation: LGPA = 0.219 (UGPA) + 0.081 (mLSAT) + 0.138 (LSAT) + 0.332 (JD-Next).

50 (median LSAT scores 161–180), Group II: those in the 50–100 range (median LSAT scores of 154–160), and Group III: those above that range (median LSAT scores of 144–153). Where students matriculated from one law school (where they joined JD-Next) to another (which may or may not have participated in JD-Next), we counted that student in the group in which they matriculated, if we could secure the first-semester grades. Table 2 shows the demographics split by school groups.

For all groups, as above, we examined both the validity of the JD-Next exam and the LSAT, including converted GRE scores. We did find a positive correlation between both tests' scores and law school grades in all these school groups.

More importantly, we also tested for incremental validity, above that provided by a base model with median LSAT, as a measure of school selectivity, and UGPA. As shown in Table 7, when examining the ΔR^2 p-values for Group I (the Top 50 schools), the JD-Next exam and LSAT both provide statistically significant improvements in predicting LGPA in all the models.

JOURNAL OF EMPIRICAL LEGAL STUDIES

TABLE 7 Sample of Group I law schools (Top 50) JD-Next course takers ($N = 93$), summary of user-determined stepwise regression analysis for the JD-Next or LSAT exams with UGPA and median LSAT predicting first semester LGPA in 2020 cohort

	First semester LGPA			
Variables entered	r	R^2	ΔR^2	ΔR^2 p
Model 1[a]				
UGPA and mLSAT	0.199	0.040	0.040	0.162
JD-Next	0.430	0.185	0.145	<0.001
Model 2[b]				
UGPA and mLSAT	0.199	0.040	0.040	0.162
LSAT	0.421	0.177	0.137	<0.001
Model 3[c]				
UGPA and mLSAT	0.199	0.040	0.040	0.162
JD-Next	0.430	0.185	0.145	<0.001
LSAT	0.502	0.252	0.067	0.006
Model 4[d]				
UGPA and mLSAT	0.199	0.040	0.040	0.162
LSAT	0.421	0.177	0.137	<0.001
JD-Next	0.502	0.252	0.075	0.004

Note: LSAT variable includes converted GRE scores. Standardized regression coefficients reported in final equations.

Abbreviations: LGPA, law school grade point average; LSAT, law school admissions test; mLSAT, median LSAT for the participants' law school; UGPA, undergraduate grade point average.

[a]Final equation: LGPA = 0.247 (UGPA) + −0.009 (mLSAT) + 0.396 (JD-Next).

[b]Final equation: LGPA = 0.308 (UGPA) + 0.053 (mLSAT) + 0.391 (LSAT).

[c]Final equation: LGPA = 0.325 (UGPA) + −0.006 (mLSAT) + 0.302 (JD-Next) + 0.290 (LSAT).

[d]Final equation: LGPA = 0.325 (UGPA) + −0.006 (mLSAT) + 0.290 (LSAT) + 0.302 (JD-Next).

For Group II schools (those in the 51–100 ranks) as shown in Table 8, the base model of UGPA and median LSAT has more predictive power in this group of schools. Nonetheless, the JD-Next score provided a statistically significant improvement in predicting LGPA in all the models. However, the LSAT's incremental predictive power is smaller and cannot be distinguished from the null in any models for the Group II schools. For this middle group of schools, the JD-Next provides clear incremental validity, even if the LSAT may not.

For Group III schools as shown in Table 9, the base model is even stronger, but the JD-Next exam again provided a statistically significant improvement in predicting LGPA in all the models. Although our point estimate for the LSAT's incremental validity is positive, it cannot be distinguished from the null for the Group III schools. For this third group of schools, the JD-Next provides clear incremental validity, even if the LSAT may not.

TABLE 8 Sample of Group II law schools (ranked 51–100) JD-Next course takers ($N = 59$), summary of user-determined stepwise regression analysis for the JD-Next or LSAT exams with UGPA and median LSAT predicting first semester LGPA in 2020 cohort

	First semester LGPA			
Variables entered	r	R^2	ΔR^2	ΔR^2 p
Model 1[a]				
UGPA and mLSAT	0.247	0.061	0.061	0.172
JD-Next	0.523	0.273	0.212	<0.001
Model 2[b]				
UGPA and mLSAT	0.247	0.061	0.061	0.172
LSAT	0.248	0.062	0.001	0.863
Model 3[c]				
UGPA and mLSAT	0.247	0.061	0.061	0.172
JD-Next	0.523	0.273	0.212	<0.001
LSAT	0.524	0.275	0.002	0.729
Model 4[d]				
UGPA and mLSAT	0.247	0.061	0.061	0.172
LSAT	0.248	0.062	0.001	0.863
JD-Next	0.524	0.275	0.213	<0.001

Note: LSAT variable includes converted GRE scores. Standardized regression coefficients reported in final equations.

Abbreviations: LGPA, law school grade point average; LSAT, law school admissions test; mLSAT, median LSAT for the participants' law school; UGPA, undergraduate grade point average.

[a]Final equation: LGPA = 0.125 (UGPA) + 0.230 (mLSAT) + 0.461 (JD-Next).

[b]Final equation: LGPA = 0.099 (UGPA) + 0.223 (mLSAT) + −0.024 (LSAT).

[c]Final equation: LGPA = 0.114 (UGPA) + 0.239 (mLSAT) + 0.463 (JD-Next) + −0.043 (LSAT).

[d]Final equation: LGPA = 0.114 (UGPA) + 0.239 (mLSAT) + −0.043 (LSAT) + 0.463 (JD-Next).

Score disparities for racial and ethnic groups

Recall that in the 2019 cohort, we oversampled under-represented student groups, and a majority of participants identified as other than White non-Hispanics. Our primary validity models do not use race or ethnicity as covariates, because we would not expect admissions officers to adjust standardized test scores for race or ethnicity. However, we were interested in the performance of the JD-Next exam for diverse populations. In particular, do students from historically marginalized or underrepresented populations tend to score lower on the exam, and is the exam's validity robust across these various groups? These questions about score disparities are important because admissions tests can impact efforts to increase diversity, equity, and inclusion in law schools. If admissions officers rely on these tests to decide which applicants to reject, and

JOURNAL OF EMPIRICAL LEGAL STUDIES

TABLE 9 Sample of Group III law schools (ranked 100+) JD-Next course takers ($N = 86$), summary of user-determined stepwise regression analysis for the JD-Next or LSAT exams with UGPA and mLSAT predicting first semester LGPA in 2020 cohort

	First semester law GPA			
Variables entered	***r***	R^2	ΔR^2	ΔR^2 ***p***
Model 1[a]				
UGPA and mLSAT	0.387	0.150	0.150	0.001
JD-Next	0.503	0.253	0.103	0.001
Model 2[b]				
UGPA and mLSAT	0.387	0.150	0.150	. 001
LSAT	0.417	0.174	0.024	0.130
Model 3[c]				
UGPA and mLSAT	0.387	0.150	0.150	0.001
JD-Next	0.503	0.253	0.103	0.001
LSAT	0.506	0.256	0.003	0.579
Model 4[d]				
UGPA and mLSAT	0.387	0.150	0.150	0.001
LSAT	0.417	0.174	0.024	0.130
JD-Next	0.506	0.256	0.082	0.004

Note: LSAT variable includes converted GRE scores. Standardized regression coefficients reported in final equations.

Abbreviations: LGPA, law school grade point average; LSAT, law school admissions test; mLSAT, median LSAT for the participants' law school; UGPA, undergraduate grade point average.

[a]Final equation: LGPA = 0.283 (UGPA) + 0.066 (mLSAT) + 0.330 (JD-Next).

[b]Final equation: LGPA = 0.323 (UGPA) + −0.029 (mLSAT) + 0.206 (LSAT).

[c]Final equation: LGPA = 0.280 (UGPA) + 0.020 (mLSAT) + 0.311 (JD-Next) + 0.076 (LSAT).

[d]Final equation: LGPA = 0.280 (UGPA) + 0.020 (mLSAT) + 0.076 (LSAT) + 0.311 (JD-Next).

lower test scores are associated with some races or ethnicities, then students with those identities are more likely to be rejected, and overall representation in law school and the legal profession is thereby reduced. Aside from simple score disparities, we are also interested in whether the exams have predictive validity in these distinct racial and ethnic groups.

For this purpose, we compare the performance of the JD-Next exam to the LSAT (including GRE scores converted to LSAT scores). In terms of the raw exam scores, we do observe differences in the scores depending on race/ethnicity for both tests. Notably, our data replicate some of the same score disparities shown by LSAC for the LSAT exam. For example, in our 2020 sample, White (non-Hispanic) test takers ($n = 145$) scored 159.51 (SD = 5.42) on the LSAT on average, while Black/African American test takers ($n = 17$) scored 149.53 (SD = 5.20) on average. This significant difference of 10 points is very similar the 10-point and 11-point differences reported in various years by LSAC, based

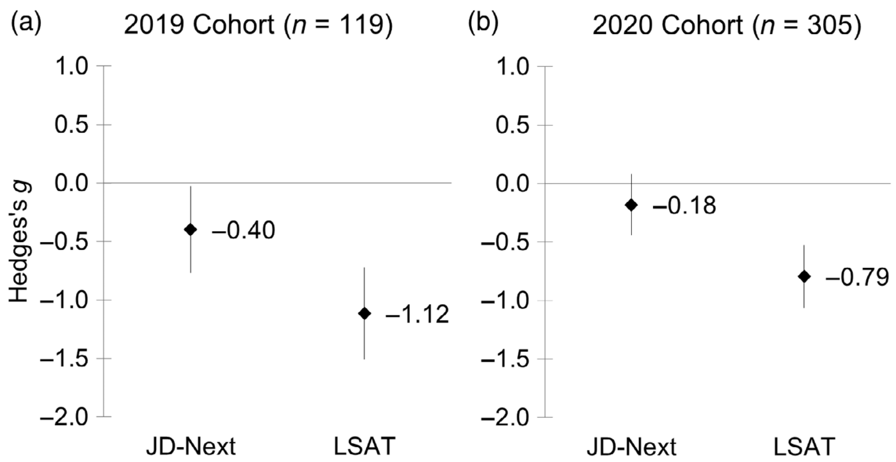

FIGURE 2 Standardized score disparities for under-represented groups (URG) on JD-Next exam and law school admissions test (LSAT) by year. 95% Confidence intervals shown for Hedges's *g*. LSAT variable includes converted GRE scores. URG pooled group includes all races and ethnicities showing significant differences from non-Hispanic Whites on at least one of the two exams (i.e., non-White Hispanic, Black, Native American and Alaska Natives, Multi-Race).

on their comprehensive census of test-taker data (Dalessandro et al., 2014). In contrast, for both exams, we find that Asian test-takers tend to perform as well, or better than White (non-Hispanic) test takers.

Given that these exams have different scales and different score distributions, it is necessary to use standardized statistics to evaluate the significance of these differences and to compare them across groups and across exams. A Cohen's *d* statistic would be the typical approach, but since the groups have different sizes, we use the Hedges's *g* statistic. Although these statistics are said to measure "effect sizes" we do not make claims of causality, and instead refer to "score disparities."

Using the Hedges *g*, for Black, Hispanic, Native American, and Multi-Race test takers, we find significant disparities on the LSAT test scores, and in every case the point estimates for the JD-Next exams trend toward smaller disparities, though not statistically distinguishable from the LSAT at these sample sizes.

Figure 2 displays Hedges's *g* statistics for the two tests, pooling the groups that showed statistically significant disparities on at least one test ("URG") and contrast them with the remainder of test-takers (White and Asian), with sample sizes shown in parentheses and 95% confidence intervals (CIs) plotted.[3] In 2019,

[3]We get similar results when using a simple contrast between Whites and non-Whites. For example, in 2020, considering White (n = 229) versus non-White (including Hispanic) (n = 78) test takers we found a significantly lower LSAT score (3.27 points on the mean, Hedges g = 0.53, p < 0.001), but only an insignificant disparity in the JD-Next score (1.28 points on the mean, Hedge's g = 0.11).

JOURNAL OF EMPIRICAL LEGAL STUDIES

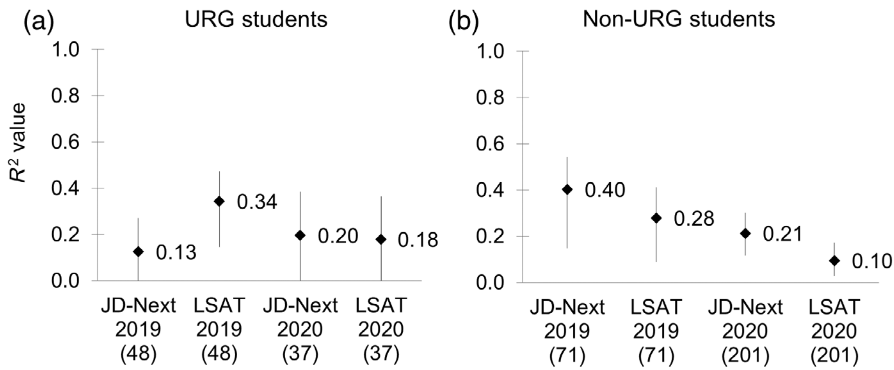

FIGURE 3 Predictive validity for underrepresented students (a) and others (b) by test and year, R^2 values, with 95% confidence interval (*N*). URG pooled group includes all races and ethnicities showing significant exam performance differences from Whites at least one of the two exams (i.e., Hispanic, Black, Native American, Multi-Race). R^2 are based on regression models with UGPA and median LSAT as covariates with JD-Next exam or law school admissions test scores, including converted GRE scores predicting LGPA. LGPA, law school grade point average; LSAT, law school admissions test; UGPA, undergraduate grade point average.

using our full sample (including the control group) to maximize power, we find significant disparities for both the JD-Next ($g = -0.40$, CI: -0.03, -0.77, $p < 0.05$) and the LSAT test scores ($g = -1.12$, CI: -0.72, -1.12, $p < 0.01$). Race/ethnicity is associated with a much smaller disparity on the JD-Next scores (between one-half and one-third the size) than the LSAT scores. In 2020 using the URG pooled group, we again see a remarkable contrast with non-URG test takers. On the LSAT, we find a difference of 4.77 points ($g = -0.79$, CI: 0.52, 1.06, $p < 0.001$). For the JD-Next, we find a nonsignificant difference of 1.99 points on the mean ($g = -0.18$, 0.08, -0.44), with a point estimate that is less than one-fourth the disparity shown on the LSAT. In short, we find substantial score disparities for the LSAT, and substantially smaller, nonsignificant disparities for the JD-Next exam.

We also examined the predictive validity of both tests for under-represented groups in both 2019 and 2020 cohorts. Figure 3 display's R^2 values for models including UGPA and school median LSAT, again grouping students ("URG") based on whether their race or ethnicity suffered significant score disparities (those other than Whites and Asians), with confidence intervals calculated in SPSS following Smithson (2001). In 2019, the JD-Next exam was a substantially better predictor for non-URG students, and the LSAT performed reasonably well for both groups. In 2020, the trends were somewhat different. The JD-Next exam had consistent predictive power across both groups, while the LSAT was somewhat weaker for non-URG students. Although we plot point estimates and 95% confidence intervals for all these groups and years, we caution that statistical power is limited for the URG students in particular.

In summary, the JD-Next exam had consistently smaller score disparities for under-represented students, and in 2020 we were unable to rule out the null hypothesis of no disparity at all. With the same statistical power, we found that the LSAT had large statistically significant disparities in scores in both years. We found positive predictive validity for both exams across both cohorts.

STRENGTHS, LIMITATIONS, AND CONCLUSIONS

At a time when many are questioning the value of admissions tests, our paper contributes to a broader literature about proximal and dynamic testing, showing how these approaches can be successfully applied to higher education, or at least professional school, admissions in the United States, where there are documented problems of diversity and representation. We found the JD-Next exam to be reliable and valid for predicting law school success, providing a significant increment in predictive power over base models.

Examining groupings of schools for validity, we did find positive correlations between test scores and law school grades, but incremental validity was more complex, and disconcerting. For law schools ranked in the Top 50, the JD-Next exam and LSAT both provided statistically significant improvements in predicting LGPA over the base model, while the JD-Next exam tended to have higher r-value point-estimates. For schools ranked outside the Top 50 (in both the 50–100 and 101+ ranks), the JD-Next exam was a statistically significant increment in predicting LGPA in all the models, but the LSAT did not provide a significant increment in predictive validity. Although statistical power is limited, these findings should be read in light of ABA Standard 503, which requires use of a valid and reliable exam as a condition of accreditation.

With regard to racial disparities, we sought to design a test that measured students' potential to learn in a supportive environment, rather than merely measuring their accumulated intellectual capital, which would be subject to the social disparities of the United States population. Accordingly, we found that the JD-Next exam tended to produce only small score disparities for under-represented groups, which were non-significant in 2020. In contrast, there were statistically significant mean differences in LSAT scores, consistent with the data reported by LSAC for their own exam (Dalessandro et al., 2014). When these findings are paired with our inability to detect significant incremental predictive value noted above, the use of the legacy exam may be problematic.

Strengths of our study include a diverse population of students, which was enriched for racial and ethnic representation in 2019, where non-Hispanic Whites represented less than half of the test-takers. Our 2019 UArizona sample and almost all of the 2020 sample has the strength of receiving transcript outcomes data directly from schools. For our key validity findings, we rely not just on raw correlations, but rather set the bar higher to test for *incremental* validity, above that provided by a base model of UGPA and median LSAT, as a measure of school selectivity. Our ability to provide parallel analyses on LSAT (with converted GRE scores) also allows comparisons for the JD-Next exam versus the "state of the art" legacy exams, and since our findings are similar to those known in the LSAT literature, we can rule out selection problems or peculiar modeling choices as a driver for our JD-Next outcomes. Moreover, the key validity findings were replicated over two distinct populations, recruited with two different sampling strategies, over 2 years, with an intervening pandemic.

Still, there are important limitations. Some of the 2019 data were self-reported, which could be infected with bias. Although our primary validity findings are based on several hundred respondents and yield highly significant results, there are fewer participants from any particular law school. Some of our null results, for example the LSAT's lack of significant incremental validity for schools in the bottom ranks, may be due to the smaller sample size in that subset.

In this paper, we focus on validity for students who were incentivized to, and in fact did, take the associated JD-Next course (including via random assignment in 2019), as the exam is designed and intended to measure potential to learn in such a course. This paper does not support validity for noncourse takers.

As a covariate in our base prediction models, we use LSAT medians as a proxy for school selectivity. In future work with larger samples, hierarchical linear modeling would be appropriate.

We emphasize that some of our analyses rely on the "LSAT" variable, but in our data, this is an index variable that includes converted GRE scores. We make no claims about either test in particular.

Future research and development would be required for the JD-Next to be used in law school admissions, with consideration of a range of issues including test security in a high-stakes environment. When given with higher stakes, the exam may have different psychometric properties.

For the more than 71,000 law school applicants each year, the JD-Next exam holds promise as a new law school admissions pathway, both to better predict success in law school and to help diversify the populations of students in law school. In this way, we hope to reduce the number of false negatives produced by current admissions practices, where capable prospective law students are denied admission and thereby excluded from the profession. In particular, we hope to reduce such exclusionary practices for groups that are already underrepresented in law schools and the profession. Moreover, our companion papers show that the JD-Next program is an effective way to prepare diverse populations of students, improving performance in law school. Aside from picking winners and losers through testing, this intentional program of development is a second way to recognize and produce capability—actually creating opportunities for successful legal study that otherwise would not exist.

Although JD-Next was philanthropically subsidized and thereby offered for free to students in the context of this research project, it does have financial costs and consumes substantial time for the students to take the associated course and submit the exam. Accordingly, it is important to consider whether JD-Next could be cost–benefit justified. On the surface, JD-Next may seem terribly inefficient since it is designed to paired with a lengthy course. That in part depends on whether it will be a complement or a substitute for extant services.

JD-Next could be used as a *complement* to this existing suite of services, for example, allowing students who are on a law school waitlist to distinguish themselves by both completing the JD-Next course and then providing a high score on the JD-Next exam. Our 2020 sample suggests that the JD-Next exam has incremental validity, even on top of a model that includes the LSAT, with UGPA and school selectivity. Imagine, for example, a Black, Hispanic, or Native American student who suffered from a score disparity under a legacy exam, but was able to demonstrate their actual abilities to study law through the greater predictive power of the JD-Next exam. For such a student, or for the law school who thereby is able to thereby confidently admit the student, the additional time and cost of the JD-Next exam may be worthwhile.

However, the greater value of the JD-Next course and exam, in terms of cost and time efficiency, may be as a *substitute* for some extant educational services. Assuming regulatory barriers can be crossed, JD-Next may simply become a third, alternative admissions pathway, which students and schools choose, alongside LSAT and, more recently, the GRE. Especially for law schools outside the Top 50, where legacy exams may provide less incremental predictive value, the JD-Next could be a substitute. Likewise, students suffering the greatest disparities on the legacy exams may find the JD-Next exam to be an attractive substitute.

In another sense the larger JD-Next program could be a substitute for other educational services, as it rebundles several key functions and makes others obsolete. Because it is not merely a test, the JD-Next program is relevant to at least three potential educational services: admissions testing, preparatory courses to help prepare applicants excel on admissions tests, and bridge programs to prepare students for law school itself. Each of these services is also costly for prospective students, in terms of time and money. In JD-Next, preparation for the admissions exam and the preparation for law school itself are offered as a single package, which potentially improves efficiency. In this way, we hope to lower the net barriers to entry for legal education and the legal profession. while improving predictions as to law school success and reducing disparities in law school admissions.

ACKNOWLEGMENTS

The research team thanks the AccessLex Institute for support of this project in the 2019 and 2020 cohorts, and the collaborative support of Marc Miller, Keith

JOURNAL OF EMPIRICAL LEGAL STUDIES

Swisher, and Rob Williams. David Klieger, Jonathan Weeks, Jonathan Steinberg, Steven Holtzman, and Dan Rodriguez provided helpful comments, in addition to those provided by three anonymous peer reviewers. For companion articles, see Cheng et al. (2021) and Buzick et al. (Buzick et al., In press).

DATA AVAILABILITY STATEMENT

Data necessary to replicate the results of this article are available upon request from the corresponding author.

REFERENCES

Amabebe, E. M. (2020). "Valid and reliable": The LSAT, ABA standard 503, and the future of law school admissions. *New York University Law Review*, *95*(6), 1860–1903.

American Bar Association. (2016). *ABA standards and rules of procedure for approval of law schools 2016–2017*. https://www.americanbar.org/content/dam/aba/publications/misc/legal_education/Standards/2016_2017_aba_standards_and_rules_of_procedure.authcheckdam.pdf

American Bar Association. (2017, November 6). *ABA law school accrediting arm proposes major change in standard for admissions tests*. https://www.americanbar.org/news/abanews/aba-news-archives/2017/11/aba_law_school_accre/

American Bar Association. (2018, May 14). *Council adopts proposal to make standardized tests optional for law schools*. https://www.americanbar.org/news/abanews/aba-news-archives/2018/05/council_adopts_propo/

Binno v. American Bar Association, 826 F.3d 338. (6th Cir. 2016).

Burton, N. W., & Wang, M. (2005). *Predicting long-term success in graduate school: A collaborative validity study*. Educational Testing Service.

Buzick, H., Findley, J., Burross, H., Charles, M., Klieger, D., & Robertson, C.T. (In press). *Increasing the pool of successful law school students: The effect of a training program on first-year law school students' grades*.

Cheng, K., Findley, J., Cimetta, A., Burross, H., Charles, M., Balser, C., Li, R., & Robertson, C. T. (2021). JD-Next: A randomized experiment of an online scalable program to prepare diverse students for law school. Arizona Legal Studies Discussion Paper (21-11).

Christensen, L. M. (2007). Legal reading and success in law school: An empirical study. *Seattle University Law Review*, *30*, 603.

Christensen, L. M. (2009). The power of skills: An empirical study of lawyering skills grades as the strongest predictor of law school success. *St. John's Law Review*, *83*, 795.

Conley, D. (2009). *Being black, living in the red: Race, wealth, and social policy in America (first)*. University of California Press.

Cortina, J. M. (1993). What is coefficient alpha? An examination of theory and applications. *Journal of Applied Psychology*, *78*, 98–104.

Cronbach, L. J. (1947). Test "reliability": Its meaning and determination. *Psychometrika*, *12*(1), 1–16.

Curtis, D. (2019). The LSAT and the reproduction rise of hierarchy. *New England Law Review*, *41*, 307–332.

Dalessandro, S. P., Anthony, L. C., & Reese, L. M. (2000). *LSAT performance with regional, gender, and racial/ethnic breakdowns: 1993–1994 through 1999–2000 testing years*. LSAT Technical Report, 100-01.

Dalessandro, S. P., Anthony, L. C., & Reese, L. M. (2014, October). *LSAT performance with regional, gender, and racial/ethnic breakdowns: 2007–2008 through 2013–2014 testing years*, LSAT Technical Report, 14-02.

Den Hartigh, R. J. R., Niessen, S. M., Frencken, W. G., & Meijer, R. R. (2018). Selection procedures in sports: Improving predictions of Athletes' future performance. *European Journal of Sport Science*, *18*(9), 1191–1198.

Ebel, R. L., & Frisbie, D. A. (1991). *Essentials of educational measurement*. Pearson College Div.

Educational Testing Service. (2020). GRE comparison tool for law schools. https://www.ets.org/gre/institutions/admissions/interpretation_resources/law_comparison_tool/

Educational Testing Service. (2021, April 14). The GRE® subject tests. https://www.ets.org/gre/institutions/about/subject/

Farley, A. N., Swoboda, C. M., Chanvisanuruk, J., McKinley, K. M., Boards, A., & Gilday, C. (2019). A deeper look at bar success: The relationship between law student success, academic performance, and student characteristics. *Journal Empirical Legal Stududies*, *622*, 605–629.

Goldman, R., Schmidt, D., Hewitt, B., & Fisher, R. (1974). Grading practices in different major fields. *American Educational Research Journal*, *11*(4), 343–357.

Guinier, L. (2015). *The Tyranny of the meritocracy: Democratizing higher education in America*. Beacon Press.

Haddon, P. A., & Post, D. W. (2006). Misuse and abuse of the LSAT: Making the case for alternative evaluative efforts and a redefinition of merit. *St. John's Law Review*, *80*, 41.

Hunsley, J., & Meyer, G. J. (2003). The incremental validity of psychological testing and assessment: Conceptual, methodological, and statistical issues. *Psychological Assessment*, *15*(4), 446–455. https://doi.org/10.1037/1040-3590.15.4.446

Jaschik, S. (2015, February 3). The Tyranny of the Meritocracy. *Inside Higher Ed*. https://www.insidehighered.com/news/2015/02/03/qa-lani-guinier-about-her-new-book-college-admissions

Jaschik, S. (2021, April 10). ABA gives law schools go-ahead to use GRE. *Inside Higher Ed*. https://www.insidehighered.com/admissions/article/2021/12/06/aba-gives-law-schools-go-ahead-use-gre

Johnson, V. E. (2003). *Grade inflation: A crisis in college education*. Springer-Verlag.

Kidder, W. C. (2000). Portia denied: Unmasking gender bias on the LSAT and its relationship to racial diversity in legal education. *Yale Journal of Law & Feminism*, *12*, 36.

Kidder, W. C. (2000). The rise of the testocracy: An essay on the LSAT, conventional wisdom, and the dismantling of diversity. *Texas Journal of Women & Law*, *9*, 147.

Klieger, D. M., Bridgeman, B., Tannenbaum, R. J., & Cline, F. A. (2016). *The validity of GRE® scores for predicting academic performance at the University of Arizona James E.* Rogers College of Law. https://online.wsj.com/public/resources/documents/gre_validitystudy_arizona.pdf

Kubiszyn, T., & Borich, G. D. (2003). *Educational testing and measuring: Classroom application and practice* (7th ed.). Wiley.

Kuncel, N. R., & Hezlett, S. A. (2007). Standardized tests predict graduate students' success. *Science*, *315*, 1080–1081. doi:10.1126/science.1136618

Law schools that accept GRE® Scores for their J.D. Programs. (2020). ETS GRE. https://www.ets.org/gre/institutions/accept/law/jd_programs/

Lidz, C. (1995). Dynamic assessment and the legacy of L.S. Vygotsky. *School Psychology International*, *16*, 243.

Marks, A. B., & Moss, S. A. (2016). What predicts law student success? A longitudinal study correlating law student applicant data and law school outcomes. *Journal of Empirical Legal Studies*, *13*(2), 205–265.

Martinez, L., West, J., & Adams, W. (2022, June 1). Matters for Notice and comment: ABA Standards 501 and 503. https://www.americanbar.org/content/dam/aba/administrative/legal_education_and_admissions_to_the_bar/council_reports_and_resolutions/may22/22-june-memo-notice-comment-501-503.pdf

Miller, M., Chemerinsky, E., Rodrigues, D. B., Morant, B., Guzman, A. T., & Farnsworth, W. (2017, June 28). *Standard 503 comments*. https://www.americanbar.org/content/dam/aba/administrative/legal_education_and_admissions_to_the_bar/council_reports_and_resolutions/

comments/20170628_comment_s503_deans_miller_chemerinsky_rodriguez_morant_guzman_farnsworth.pdf
Niessen, S. M., Meijer, R. R., & Tendeiro, J. N. (2016, April 13). Predicting performance in higher education using proximal predictors. *PLoS One*, *11*(4), 1–14.
Niessen, S. M., Meijer, R. R., & Tendeiro, J. N. (2018, June 11). Admission testing for higher education: A multi-cohort study on the validity of high-Fidelity curriculum-sampling tests. *PLoS One*, *13*(6), 1–21.
Organ, J. M. (2017). Net tuition trends by LSAT category from 2010 to 2014 with thoughts on variable return on investment. *Journal of Legal Education*, *67*, 51.
Randazzo, S. (2016, February 22). Move over LSAT, here comes the GRE. *Wall Street Journal Law Blog*. https://www.wsj.com/articles/BL-LB-53187
Reardon, S. F., Kalogrides, D., & Shores, K. (2019). The geography of racial/ethnic test score gaps. *American Journal of Sociology*, *124*(4), 1164–1221.
Robertson, C. T., Miller, M. L., Williams, R. A., & Pierre, J. K. (2022, August 16). It's time to repeal the ABA's law school testing mandate. Bloomberg law. https://news.bloomberglaw.com/us-law-week/its-time-to-repeal-the-abas-law-school-testing-mandate
Roth, P. L., Bobko, P., & McFarland, L. A. (2005). A meta-analysis of work sample test validity: Updating and integrating some classic literature. *Personnel Psychology*, *58*, 1009–1037.
Sampson, C., & Boyer, P. G. (2001). GRE scores as predictors of minority Students' success in graduate study: An argument for change. *College Student Journal*, *35*, 271.
Schmidt, F. L., & Hunter, J. E. (1998). The validity and utility of selection methods in personnel psychology: Practical and theoretical implications of 85 years of research findings. *Psychological Bulletin*, *124*, 262–274.
Shultz, M. M., & Zedeck, S. (2011). Predicting law effectiveness: Broadening the basis for law school admission decisions. *Law & Social Inquiry*, *36*, 620–661.
Shultz, M. M., & Zedeck, S. (2012). Admission to law school: New measures. *Educational Psychologist*, *47*(1), 51–65.
Smithson, M. (2001). Correct confidence intervals for various regression effect sizes and parameters: The importance of noncentral distributions in computing intervals. *Educational and Psychological Measurement*, *61*, 605–632.
Sternberg, R. J., & Grigorenko, E. L. (2020). *Dynamic testing: The nature and measurement of learning potential*. Cambridge University Press.
Tavakol, M., & Dennick, R. (2011). Making sense of Cronbach's alpha. *International Journal of Medical Education*, *2*, 53–55.
Taylor, A. N. (2019). The marginalization of black aspiring lawyers. *Florida International University Law Review*, *13*, 489.
Taylor, A. N., Scott, J. M., & Jackson, J. (2021, March 24). *It's not where you start, It's how you finish: Predicting law school and Bar success. AccessLex*. https://www.accesslex.org/research-and-data-tools-and-resources/its-not-where-you-start-its-how-you-finish-predicting-law
Thomas, D. A. (2003). Predicting law school academic performance from LSAT scores and undergraduate grade point averages: A comprehensive study. *Arizona State Law Journal*, *35*, 1007.
Tucker, G. C. (2016, October 26). *Putting pre-employment tryouts to the test*. Society for Human Resource Management. https://www.shrm.org/hr-today/news/hr-magazine/1116/pages/putting-pre-employment-tryouts-to-the-test.aspx
Whitford, W. C. (2017). Law school-administered financial aid. *Journal Legal Education*, *67*, 4.
Wightman, L. F. (2000). *Beyond FYA: Analysis of the utility of LSAT scores and UGPA for predicting academic success in law school*. Law School Admission Council.
Zirkel, P. A. (2019). Law school scholarships: Hooded Robin' of non-privileged students. *Journal of Legal Education*, *48*, 239.
Zwick, R. (2013). *Rethinking the SAT: The future of standardized testing in university admissions*. Routledge.

How to cite this article: Findley, J., Cimetta, A., Burross, H. L., Cheng, K. C., Charles, M., Balser, C., Li, R., & Robertson, C. (2023). JD-Next: A valid and reliable tool to predict diverse students' success in law school. *Journal of Empirical Legal Studies*, 1–32. https://doi.org/10.1111/jels.12342