Measuring the Power of Learning.™

Research Report

ETS RR–18-26

The Validity of *GRE*® General Test Scores for Predicting Academic Performance at U.S. Law Schools

David M. Klieger

Brent Bridgeman

Richard J. Tannenbaum

Frederick A. Cline

Margarita Olivera-Aguilar

December 2018

ETS Research Report Series

Since its 1947 founding, ETS has conducted and disseminated scientific research to support its products and services, and to advance the measurement and education fields. In keeping with these goals, ETS is committed to making its research freely available to the professional community and to the general public. Published accounts of ETS research, including papers in the ETS Research Report series, undergo a formal peer-review process by ETS staff to ensure that they meet established scientific and professional standards. All such ETS-conducted peer reviews are in addition to any reviews that outside organizations may provide as part of their own publication processes. Peer review notwithstanding, the positions expressed in the ETS Research Report series and other published accounts of ETS research are those of the authors and not necessarily those of the Officers and Trustees of Educational Testing Service.

The Daniel Eignor Editorship is named in honor of Dr. Daniel R. Eignor, who from 2001 until 2011 served the Research and Development division as Editor for the ETS Research Report series. The Eignor Editorship has been created to recognize the pivotal leadership role that Dr. Eignor played in the research publication process at ETS.

ETS Research Report Series ISSN 2330-8516

RESEARCH REPORT

The Validity of *GRE*® General Test Scores for Predicting Academic Performance at U.S. Law Schools

David M. Klieger, Brent Bridgeman, Richard J. Tannenbaum, Frederick A. Cline, & Margarita Olivera-Aguilar

Educational Testing Service, Princeton, NJ

Educational Testing Service (ETS), working with 21 U.S. law schools, evaluated the predictive validity of the *GRE*® General Test using a sample of 1,587 current and graduated law students. Results indicated that the GRE is a strong, generalizably valid predictor of first-year law school grades and that it provides useful information even when undergraduate grade point average already is available to predict those grades. This report also reviews the reliability of the GRE General Test that had been determined in prior research.

Keywords *GRE*®; LSAT; first-year law school grades; law school admissions testing; predictor; validity

doi:10.1002/ets2.12213

The *GRE*® General Test is a large-scale assessment used to help make admissions decisions for applicants to graduate school and master's of business administration (MBA) programs. The GRE test provides decision makers with information about the verbal reasoning, quantitative reasoning, and analytical writing skills of applicants to graduate and professional schools. Given the pervasive importance of these skills to success in graduate and professional education (Kuncel & Hezlett, 2007), previous research has demonstrated empirically that GRE scores are indicative of success in PhD, MA, and MS programs across disciplines (Klieger, Cline, Holtzman, Minsky, & Lorenz, 2014; Kuncel, Hezlett, & Ones, 2001; Kuncel, Wee, Serafin, & Hezlett, 2010); in MBA programs (Young, Klieger, Bochenek, Li, & Cline, 2014); and in veterinary medical colleges (Powers, 2004). In order to determine the validity of using GRE scores for forecasting success at U.S. law schools, we partnered with 21 U.S. law schools to gather data from enrolled and graduated law students.

This predictive validity study began as an effort to show that the GRE test potentially can provide several benefits to law schools, their applicants, and the legal profession.[1] Verbal reasoning and analytical writing skills are fundamental to successful performance as a law student (and future lawyer). Law students are required to make sense of complex subject matter from ongoing and sizeable assigned readings. Furthermore, law students must express their understanding of legal issues on written examinations (the major or sole basis for course grades), in legal writing classes, and sometimes in clinical work representing clients. Moreover, the American Bar Association (ABA) advises that prospective law students should have an understanding of "basic mathematical and financial skills, such as an understanding of basic pre-calculus mathematics and an ability to analyze financial data" (ABA, 2017b, Background Knowledge section, bullet 2). Experts on lawyers' performance have observed that attorneys require and utilize quantitative reasoning skills in order to represent clients successfully in business transactional work (Gabaldon, 2014). A potential lack of quantitative reasoning skills among lawyers is increasingly problematic as data-driven, scientific, and technological matters more frequently become the subject of litigation (Weiss, 2013). Martha Minow, the dean of Harvard Law School, observed:

> [G]iven the promise of the revolutions in biology, computer science, and engineering, law needs students with science, technology, engineering and math backgrounds. For these students, international students, multidisciplinary scholars, and joint-degree students, the GRE is a familiar and accessible test, and using it is a great way to reach candidates not only for law school, but for tackling the issues and opportunities society will be facing. (In pilot program, 2017, para. 4).

Corresponding author: D. M. Klieger, E-mail: dklieger@ets.org

As this statement indicates, the GRE test could help expand access to legal education beyond the traditional pre–law degree fields. There are potential law school applicants who have either completed or are considering many science, technology, engineering, and mathematics (STEM) and non-STEM graduate and professional programs that require or recommend the GRE test. Prospective law students may lack the financial resources or time to prepare for and take more than one admissions test, or they may be considering a dual degree that would already require them to take the GRE test. In addition, the GRE test is administered by computer year-round in more than 1,000 testing centers located across the United States and in more than 160 countries. More people take the GRE test than any other graduate or professional school admissions test, with more than half a million test takers per annum over the past few years (see Educational Testing Service [ETS], 2014b). By accepting GRE test scores, law schools potentially can encourage applications from students with a wider range of academic backgrounds.

Given these potential benefits to applicants, institutions, and the legal profession, a remaining prerequisite to the use and interpretation of GRE scores for law school admissions is confirmation that GRE scores are a valid and reliable measure of an applicant's capabilities. In addition to abiding by the professional validation standards in educational testing (American Educational Research Association [AERA], American Psychological Association [APA], & National Council on Measurement in Education [NCME], 2014; ETS, 2014a), this GRE test validity research for law school also is responsive to the ABA Section of Legal Education Standard 503 that was in effect when this study was conducted (ABA, 2014). Standard 503 required that for a law school to be accredited by the ABA, it must require its applicants 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, 2014, p. 33). At the time of this study, most applicants to ABA-accredited law schools submitted scores from the Law School Admission Test (LSAT).

This validation study uses scores from only the current version of the GRE test, which was launched in August 2011. The GRE test provides test takers and institutions with three scores, one for each of three sections: verbal reasoning (GRE-V), quantitative reasoning (GRE-Q), and analytical writing (GRE-AW). The measures strongly emphasize complex verbal and quantitative reasoning skills via text-based materials, real-life scenarios, and data interpretation (Briel & Michel, 2014). Wendler and Bridgeman (2014) provided detailed information about the changes to the GRE test that became effective in August 2011. What follows is a summary. Based heavily on reading passages, GRE-V evaluates reading comprehension as well as verbal and analytical reasoning skills. Test takers are required to analyze and evaluate written material. GRE-Q measures problem-solving skills and includes questions assessing financial literacy, data interpretation, and arithmetic abilities. Quantitative reasoning questions focus on real-life scenarios and emphasize data interpretation. To reduce emphasis on pure computation skills, the quantitative reasoning section provides an on-screen calculator. GRE-AW measures critical thinking and analytical writing skills. Test takers are asked to demonstrate proficiency in expressing complex ideas clearly. To address the possibility of a test taker using a prepared (or "canned") response, GRE-AW requires focused responses to specific questions.

In August 2011, the way in which the GRE test was delivered and scored changed. The GRE is now given as a computerized adaptive multistage test (MST), which supports increased security and continuous testing year-round and allows test takers to go back and change their responses. Yan, von Davier, and Lewis (2014) provided more information about MST design. Scaled scores on the multiple-choice GRE-V and the GRE-Q sections are reported on a scale that ranges from 130 to 170 in 1-point increments. This scale range was chosen to avoid confusion with the pre-August 2011 GRE-V and the GRE-Q score scales that ranged from 200 to 800. Given the very limited nature of the change to GRE-AW and its two writing tasks (analyze an issue and analyze an argument) in August 2011, the 0–6 scale with half-point increments was retained from the previous version of the GRE test.

The general research question in this study is the following: To what extent are GRE-V, GRE-Q, and GRE-AW scores valid and reliable measures of law school applicants' capabilities? As discussed in more detail below, we used first-year law school grades (LGPA) as criterion measures of applicants' capabilities. Our research focused more on the question of validity than reliability: Mathematically, reliability sets the ceiling for criterion-related validity (Furr & Bacharach, 2008; Nunnally, 1978). In answering the validity research question we found considerable validity evidence, so consequently there must be a sufficient level of reliability to interpret and use GRE scores to predict LGPA. More specific validity research questions are as follows:

1. What is the predictive validity of each GRE section?
2. What is the predictive validity when scores for the different GRE sections are combined?

3. What is the predictive validity when GRE section scores are considered in conjunction with undergraduate grade point average (UGPA)? This question investigates the extent to which the GRE sections improve predictions beyond what can be predicted from UGPA alone.

Previous Research on the Validity of GRE Scores for Predicting Graduate and Professional School Performance

More than 1,500 studies have examined the validity of using GRE scores to predict performance in graduate and professional schools. Below, we summarize three large studies that evidence GRE predictive validity across a range of graduate and professional school programs.

Kuncel et al. (2001) meta-analyzed GRE predictive validity for more than 82,000 students across more than 1,700 samples. They found that GRE-V and GRE-Q scores were correlated with several graduate school outcomes in general and several discipline areas (humanities, social sciences, life sciences, and mathematics/physical sciences) in particular.[2] Validity coefficients (i.e., correlations between test scores and the predicted outcome) adjusted for range restriction and/or unreliability generally were in the .30s – .50s for predicting grades. Overall and for specific discipline areas, GRE-V and GRE-Q almost always predicted graduate grade point average (GGPA) and first-year GGPA at least as strongly as did UGPA.

Klieger et al. (2014) studied the predictive validity of GRE scores for 25,356 students across 28 master's and 26 doctoral program areas in 10 Florida public universities. Overall, for both master's and PhD students, and by discipline area, each of the three GRE measures predicted cumulative GGPA.

Young et al. (2014) validated the use of GRE scores in MBA program admissions. Their study included data for 480 MBA students across 12 MBA programs. For predicting first semester MBA GPA and cumulative MBA GPA, GRE-Q was most predictive, followed by GRE-V and GRE-AW. Additional analyses revealed that the three GRE sections added a great deal of predictive value over and above the use of just UGPA, and UGPA added very little predictive value over and above the use of just GRE scores.

Research Methods

Samples

The validity analyses in this study are based on data collected for 1,587 law students from 21 U.S. law schools. In 2015, the University of Arizona James E. Rogers College of Law asked ETS if it would examine the validity of the GRE test for use in law school admissions. Three additional law schools that learned about ETS's work with the University of Arizona made similar requests. Those four law schools formed the start of our sample. We considered that there might be further interest from additional schools given that the ABA (2017a) accredited 205 law schools. Conducting a separate study for up to 205 schools was deemed both inefficient and unnecessary for determining whether use of GRE scores was valid and reliable for additional law schools. Under certain assumptions, which we discuss later, validity generalization methods allow one to generalize validity based on a sample of schools from the larger school population of ultimate interest (see AERA et al., 2014; Murphy, 2003; Schmidt & Hunter, 2015).

The primary objective of this research was to evaluate how well GRE scores predict LGPA across a sample of law schools and to examine how well the results generalize across ABA-approved law schools. Although a true random sample was not practically achievable, we created a sampling design so that we would obtain LGPA, GRE, LSAT, and UGPA data for students from a set of law schools that was reasonably representative of ABA-approved law schools in the United States. Based on law school selectivity (using median LSAT and UGPA of entering students; ABA, 2017c), we created three groups: most selective, moderately selective, and least selective. We also sought school diversity in terms of law school geographic region, size of the student body, and public or private status. Within each school selectivity group, we created subgroups for geographic region: northeast, south and southwest, midwest, and west. Within this structure, we identified law schools and chose to recruit schools that varied in size and public versus private status. For recruitment, we randomly chose which ABA-approved schools to contact within each selectivity – geographic group. However, over the course of this study, many law schools reached out to us with interest in participating in the national study, and we worked with an interested school if it satisfied an unmet need of the school sampling design (see Appendix A for a list of all participating law schools).

The sample of students for this report matriculated between the 2010–2011 and 2016–2017 academic years (inclusive) and had each of the following: GRE scores, LSAT scores, UGPA, and LGPA. Data were combined across matriculating years. Law schools' records contain students' LSAT, UGPA, and LGPA information but usually do not include GRE scores. For this study, all GRE scores used in analyses were from the current version of the GRE (which became operational in August 2011), and they were obtained via two processes. Ten institutions worked with ETS to identify students who took the GRE test during or after August 2011. The large majority of identified students had taken the GRE test prior to their matriculation into law school, and a few had taken the GRE test after completing law school.

Law schools with fewer students tended not to have sufficient numbers of students with pre-existing GRE scores. To accommodate these 11 schools, we included a second data collection method. For these schools, current (at the time of this study) law students took the GRE test free-of-charge at a special testing administration held at their respective law schools or off-campus at official GRE testing centers. Law schools offered reasonable compensation to students (usually between $75 and $100 per student) who took the GRE test as part of this study. Typically, part of the compensation was for taking the test; the other part was for demonstrating a good faith effort. For example, some schools encouraged a good faith effort by informing students that they would receive additional compensation if their GRE test performance was no lower than 20 percentile points below their LSAT performance. At some of the 11 institutions that tested students, additional students with prior GRE scores were identified in order to increase sample size. Because of the possibility that the data collection method (obtaining operational scores from students who had taken the GRE prior to this study versus GRE testing as part of this study) might result in two subpopulations differing from each other in terms of skill levels and/or levels of test-taking motivation, we conducted analyses to determine whether validity coefficients (i.e., correlations between the GRE test and LGPA) were comparable for the two data collection approaches. For the 812 students who took the test prior to the study, the adjusted GRE V + Q validity coefficient (with accompanying 95% confidence intervals in square brackets) was .51 [.43, .58] when compared with .55 [.43, .67] for the 775 students who tested for this study. Corresponding values for the LSAT were .53 [.46, .59] for students who tested prior to the study as compared to .57 [.42, .71] for students who tested for the study. Therefore, we did not consider the data collection method to be a substantive moderator.

Data were provided in a manner that complied with rules and regulations designed to protect human subjects and their privacy. UGPA, LSAT score, and LGPA data provided to ETS never were associated with student names in any working data files, and law schools were prevented from seeing or having access to the students' GRE scores. All law schools participating in this research contractually were required to obtain an exemption or approval from their institution's Institutional Review Board (IRB). All data sharing was arranged to meet the requirements of the Family Educational Rights and Privacy Act (FERPA).

Prior empirical research on post-secondary standardized admissions testing indicates that institutional selectivity might moderate validity (see, e.g., Allen & Robbins, 2010; Kobrin, Patterson, Shaw, Mattern, & Barbuti, 2008; Mattern & Patterson, 2011a, 2011b, 2011c). Therefore, we created subgroupings of schools based on schools' LSAT medians from Standard 509 filings reported to the ABA (2017c) for the 2016 entering classes at each of the 21 participating law schools. We divided schools into three groups based on the median LSAT scores of their enrolled students. Group I consists of schools that are among the top 14 ABA-approved law schools in terms of median LSAT scores. Schools in Group III have median LSAT scores below those of the top 50 ABA-approved schools. The median LSAT scores for Group II schools fall in between those of Group I and Group III.

For each of the three groups, separate validity analyses were conducted, and results are reported below. Six public and 15 private law schools are represented among the 21 schools. Most U.S. law schools are private (see ABA, 2017c). As can be seen in Table 1, schools from all three selectivity groups are located across various regions of the United States, range in student population size, and vary in terms of median UGPAs and median LSAT scores. Figure 1 provides a graphical depiction of the geographical distribution of the schools. Table 2 provides additional information about student sample statistics. As discussed in greater detail in a later section, GRE composites were created through z-score transformations (i.e., standardizing so that all sections are on the same scale) of the GRE-V, GRE-Q, and GRE-AW scores. After each z-score computation, the z-scores were added up.

Measures

We report findings for the following measures: GRE, LGPA, LSAT, and UGPA. LGPA, the criterion, is a critical measure of success because of the strong relationship between first-year law school grades and program completion (Wightman,

Table 1 Participating Law Schools' Characteristics, by Group and Total

	Group I	Group II	Group III	All groups
Number of schools	6	8	7	21
School type				
Public	2	2	2	6
Private	4	6	5	15
School region				
Northeast	1	1	2	4
South and southwest	2	3	1	6
Midwest	2	1	1	4
West	1	3	3	7
Number of matriculants				
Mean	274	186	169	205
Range (by school)	215–391	102–261	65–264	65–391
Reported median UGPA[a]				
Mean	3.76	3.68	3.25	3.56
Range (by school)	3.67–3.86	3.52–3.80	2.89–3.58	2.89–3.86
Reported median LSAT[a]				
Mean	169	163	152	161
Range (by school)	167–172	161–165	14–158	143–172

Note. LSAT = Law School Admission Test; UGPA = undergraduate grade point average.
[a]Each law school reported for itself a median UGPA and LSAT score for its 2016 entire entering class. The mean of medians is an average of reported values for each group and across the three groups, and the range is the highest and lowest of the reported medians for each group and across the three groups. Adapted from "Section of Legal Education: ABA Required Disclosures. 2016," by American Bar Association, 2017a, 2017b, 2017c. Retrieved from http://www.abarequireddisclosures.org/Disclosure509.aspx. Copyright 2017 by the American Bar Association.

2000), law review selection, and employment opportunities (Wecker, 2012). In addition, LGPA is a measure of law school student success that commonly is used to validate use of LSAT scores (see, e.g., Anthony, Dalessandro, & Trierweiler, 2016). Therefore, this report uses first-year LGPA as the achievement measure of students' academic performance.

Because all students in our sample took the LSAT, which at the time was the entrance examination most commonly used by ABA-accredited law schools, we report findings for the LSAT in the tables in this paper in order to better contextualize GRE findings. Because the LSAT reports only a single aggregate score that reflects performance on its logical reasoning, reading comprehension, and analytical reasoning sections, some of the information that we report is for composites of the GRE sections where each constituent GRE section is weighted the same after z-score transformation of the section scores. Z-scores below the mean are negative in sign, and those above the mean are positive in sign. Given that there was no way to calculate separate section scores for the LSAT, the creation of GRE score composites in this study should be viewed as necessary if one wishes to contextualize GRE with an LSAT approach to the weighting of composite sections.

For several analyses, we treated GRE sections as distinct, and in several cases we allowed each GRE section to take on its own optimal regression weight so that one could see the fuller potential of the GRE. In instances where GRE scores for an individual were available from more than one test administration, the scores used in the analyses were selected based on the method the institution used when presented with multiple LSAT scores, which in most instances was highest score, but in some was average score. We report findings also for UGPA alone and used together with the GRE or LSAT. Guidance on ABA Standard 503 (ABA, 2015) refers to the validity of LSAT scores when used in conjunction with UGPA. Also, we wanted to examine the validity of GRE scores as predictors of LGPA in a context where an admissions officer or committee already possessed UGPA to make admissions decisions.

Statistical Analysis Methods

A valid test must be reliable, because mathematically reliability sets the ceiling for criterion-related validity (Furr & Bacharach, 2008; Nunnally, 1978). Because GRE-Q and GRE-V are computerized adaptive MSTs (see Robin & Steffen, 2014), we used an item response theory–based method to calculate internal consistency reliability for those sections (see van Rijn, 2016).[3]

Figure 1 Geographic diversity of the 21 law schools participating in the study. Each pin represents a school. Where pins greatly overlap, a different icon was used.

For GRE-AW, we used classical test theory to calculate internal consistency reliability (see Robin & Kim, 2014). The reliability for GRE-AW is estimated from the correlation between two writing tasks (analyze an issue and analyze an argument) that are intended to measure slightly different aspects of analytical writing skills. Therefore, the reliability estimate provided for GRE-AW scores is a lower bound on the actual reliability of the scores on the combined tasks, and it should not be directly compared to the reliability estimates for GRE-V and GRE-Q scores. We found no practically significant differences in GRE reliability estimates across U.S. regions. Therefore, we report reliability numbers for a national test taker group regardless of the regions from which participating law schools primarily have been drawing their applicants. The Law School Admission Council (LSAC) has reported estimates of LSAT reliability but did not provide details of how they were calculated (see LSAC, 2017).

In analyzing validity, we sought convergence from four different analysis methods: (a) correlation and ordinary least squares regression, (b) Schmidt and Hunter (2015) meta-analysis, (c) contingency tables reporting percentages of GRE test takers falling into the top versus bottom third of GRE scores and LGPA, and (d) hierarchical linear modeling (HLM). Some analyses employed correlation and regression methods, with appropriate adjustments made to R^2 values[4] and to compensate for range restriction.[5] Because the validity coefficients necessarily are based only on data from students who were accepted and then enrolled, they reflect relationships only within this enrolled group. To estimate what the correlation would have been for the entire applicant pool, we adjusted the correlation coefficients to account for range restriction. We assumed that the applicant pool consisted not just of those who actually applied to a law school but rather anyone who expressed an interest in attending any law school by taking the LSAT.

We calculated adjusted correlation coefficients for both the LSAT and GRE using adjustment procedures that are common in predictive validity studies. LSAT scores were adjusted for direct range restriction via Thorndike Case II, and GRE scores were adjusted for indirect range restriction via Thorndike Case III (Thorndike, 1949).[6] These adjustments require estimates of the standard deviation of the LSAT scores in the applicant population. For schools that drew applicants primarily from a limited geographic region, we used the appropriate regional LSAT standard deviations as reported in Dalessandro, Anthony, and Reese (2014), and for schools that drew applicants from a broader national population, we used the national standard deviations from the same source. These adjusted correlation coefficients are imperfect estimates because they assume that applicants were selected only on LSAT scores and that the regression line (and variation around the regression line) in the selected group would also hold in the group not selected. Even though these assumptions cannot be empirically confirmed, adjusted correlations are likely to be considerably more accurate than unadjusted ones (see Schmidt & Hunter, 2015). For each law school in the sample we weighted both the unadjusted and adjusted

Table 2 Student Sample Statistics by Group and Total

	Group I	Group II	Group III	All groups
Number of schools	6	8	7	21
N (sample size)				
Total no. of students	569	476	542	1,587
Mean school size	95	60	77	76
Range (by school)	69 – 143	34 – 88	67 – 102	34 – 143
GRE-V				
Mean (*SD*)	163.4 (4.8)	159.9 (5.5)	151.8 (7.1)	158.4 (7.7)
Range (by school)	161.7 – 164.8	158.2 – 162.8	146.4 – 156.0	146.4 – 164.8
GRE-Q				
Mean (*SD*)	157.6 (6.5)	154.2 (6.9)	146.0 (6.3)	152.6 (8.2)
Range (by school)	155.3 – 160.5	151.2 – 157.8	141.5 – 149.1	141.5 – 160.5
GRE-AW				
Mean (*SD*)	4.66 (0.68)	4.49 (0.70)	3.98 (0.83)	4.38 (0.79)
Range (by school)	4.44 – 4.80	4.38 – 4.75	3.55 – 4.40	3.55 – 4.80
GRE V + Q composite[a]				
Mean (*SD*)	2.34 (1.24)	1.44 (1.38)	−0.66 (1.53)	1.05 (1.89)
Range (by school)	1.83 – 2.90	0.85 – 2.16	−1.94 – 0.26	−1.94 – 2.90
GRE V + Q + AW composite[a]				
Mean (*SD*)	3.40 (1.68)	2.28 (1.86)	−0.49 (2.24)	1.73 (2.55)
Range (by school)	2.59 – 4.02	1.55 – 3.28	−2.33 – 0.68	−2.33 – 4.02
UGPA				
Mean (*SD*)	3.64 (0.31)	3.54 (0.34)	3.21 (0.43)	3.46 (0.41)
Range (by school)	3.45 – 3.73	3.46 – 3.65	2.94 – 3.36	2.94 – 3.73
LSAT				
Mean (*SD*)	166.8 (6.1)	161.5 (5.4)	152.5 (6.7)	160.3 (8.6)
Range (by school)	163.3 – 170.0	159.9 – 164.7	145.8 – 158.2	145.8 – 170.0

Note. GRE-AW = GRE analytical writing section; GRE-Q = GRE quantitative reasoning section; GRE-V = GRE verbal reasoning section; LSAT = Law School Admission Test; UGPA = undergraduate grade point average.

[a]The GRE composites were created through z-score transformation of the GRE-V, GRE-Q, and GRE-AW scores. After the z-score computation, the three z-scores were summed within each school, and these sums then were weighted by sample size and averaged over all of the schools in a group. Values for composites can be negative in sign, because the z-scores have a mean of zero, and roughly half of them in a reference population will be negative.

correlations by the sample size in the school and then computed weighted averages for both unadjusted and adjusted correlations across the 21 schools.

Any statistical estimate necessarily reflects some degree of uncertainty that can be reported. We used statistical bootstrapping methods for the correlations to estimate sampling error (Efron, 1979).[7] For the Schmidt and Hunter (2015) meta-analytic methods, we used confidence intervals to achieve the same purpose as well as credibility intervals to characterize the variability of the true correlations across institutions. Additional details appear in the Results and Discussion section and in Appendix B.

Because the foregoing statistical methods use correlations that are not always easily understood as validity indicators, we also determined that the percentage of students in the top third versus the percentage of students in the bottom third of GRE scores corresponds to the student being in the top third versus bottom third of LGPA (based on the methods of Bridgeman, Burton, & Cline, 2009). Last, we conducted HLM (Raudenbush & Bryk, 2002) to directly model possible reasons for any observed validity differences among law schools as well as possible effects of institutional differences in selectivity.

Results and Discussion

Reliability

Because this is a national study, we report reliability in Table 3 for national GRE and LSAT test-taking populations. The reliability of GRE test scores is high with values ranging from .79 (GRE-AW) to .96 (GRE V + Q). As mentioned previously,

Table 3 Reliabilities and Observed Intercorrelations for the Full Sample ($N = 1{,}587$)

Measure	GRE V + Q + AW	GRE V + Q	GRE-V	GRE-Q	GRE- AW	UGPA	LSAT
GRE V + Q + AW	.94	.94	.88	.84	.77	.40	.80
GRE V + Q		.96	.91	.92	.49	.39	.85
GRE-V			.92	.68	.51	.38	.81
GRE-Q				.93	.39	.33	.75
GRE-AW					.79	.30	.44
UGPA						—	.37
LSAT							.90 – .95

Note. GRE-AW = GRE analytical writing section; GRE-Q = GRE quantitative reasoning section; GRE-V = GRE verbal reasoning section; LSAT = Law School Admission Test; UGPA = undergraduate grade point average. The values in the diagonal (shaded cells) represent internal consistency reliability estimates for U.S. national testing samples. Estimated reliability values for the LSAT adapted from "Law School Admission Test (LSAT): LSAT Scores as Predictors of Law School Performance," by LSAC, 2017. Retrieved from https://www.lsac.org/jd/lsat/your-score/law-school-performance. Copyright 2017 by LSAC. The reported reliability values for GRE sections and composites are based on the GRE testing population sample of 362,844 test takers who tested in the United States in 2015. Furthermore, we note that the reliability for GRE-AW is estimated from the correlation between two writing tasks (analyze an issue and analyze an argument) that are intended to measure slightly different aspects of analytical writing skills. The values in the off-diagonal (unshaded cells) represent observed correlations among LGPA, GRE, UGPA, and the LSAT for the 1,587 students across the 21 law schools participating in this national study. No distinction was made for the law school attended. All of these correlations are statistically significant ($p < .001$).

the reliability estimate provided for GRE-AW scores is a lower bound on the actual reliability of the scores on the combined GRE-AW tasks, and it should not be directly compared to the reliability estimates for the GRE-V and GRE-Q scores. The GRE V + Q + AW and GRE V + Q composite reliability estimates (.94 and .96, respectively) are comparable to LSAT reliability estimates (.90 to .95).

Correlation and Regression-Based Evidence of Validity

Table 3 provides observed correlations[8] among all of the predictors. In Table 3, all students, regardless of school, were treated as a single group. Schools can differ in student selectivity, distributions of law school grades, and other characteristics that may impact validation of test use, so for determining validity, it is better practice to treat each school as a separate unit and then combine across schools. We created sample-size-weighted averages for the three selectivity groups described above as well as for all groups combined. These correlation coefficients are in Table 4. The bottom half of Table 4 presents the coefficients adjusted for range restriction.

Even with the relatively large sample size in "All Groups," the true population[9] correlation coefficients could differ somewhat from the estimates in Table 4. For the GRE V + Q + AW adjusted correlation coefficient, the confidence interval as estimated from the bootstrap ranged from .47 to .57, and for the LSAT, the interval was .47 to .60. Therefore, for both the GRE and LSAT, we can be reasonably confident that the true mean correlation of test scores and grades is at least .47.

In Table 5, we report test validity findings that consider UGPA. Guidance on ABA Standard 503 (ABA, 2015) refers to the validity of LSAT scores when used in conjunction with UGPA. Also, we want to convey the incremental validity of GRE scores in a context where an admissions officer or committee already possesses UGPA to make admissions decisions. Table 5 shows the validity of UGPA together with ("+") GRE (and, for context, with LSAT), and it indicates the extent to which GRE scores and LSAT scores are indicative of law school grades beyond what could have been achieved with UGPA alone. This incremental validity of the test scores (ΔR^2) is an R^2 value in the GPA row (for the school selectivity group of interest or for all groups) subtracted from the R^2 value in the same column for a "+" row of the test of interest.

The multiple correlations (R) and coefficients of determination (R^2, the percentages of variance in the criterion accounted for by the indicators) were computed separately in each school, weighted by the sample size in each school, and averaged. For the "separately entered" row, each GRE score was entered into the regression equation separately, which produced the optimal weight on each score. For the equally weighted V + Q + AW row, a single GRE score was entered as the sum of the z-scores for GRE-V, GRE-Q, and GRE-AW. For each group of schools, and for All Groups, the values in the "+" rows for the GRE test are notably larger than those in the UGPA row. These differences show that GRE scores

Table 4 Observed and Adjusted Correlation Coefficients of First-Year Law School Grades (LGPA) with Undergraduate Grade Point Average (UGPA) and Admission Tests

	Group I	Group II	Group III	All groups
Number of schools	6	8	7	21
Number of students	569	476	542	1, 587
Observed coefficients (averaged) of first year LGPA with:				
GRE V + Q + AW	.32	.34	.38	.35
GRE V + Q	.34	.29	.36	.33
GRE-V	.27	.25	.31	.28
GRE-Q	.29	.26	.30	.28
GRE-AW	.14	.25	.26	.21
UGPA	.22	.17	.27	.22
LSAT	.30	.32	.37	.33
Adjusted coefficients (averaged) of first year LGPA with:				
GRE V + Q + AW	.49	.54	.57	.53
GRE V + Q	.52	.53	.58	.54
GRE-V	.46	.49	.54	.50
GRE-Q	.47	.49	.51	.49
GRE-AW	.17	.27	.39	.28
UGPA	—	—	—	—
LSAT	.51	.55	.59	.55

Note. GRE-AW = GRE analytical writing section; GRE-Q = GRE quantitative reasoning section; GRE-V = GRE verbal reasoning section; LSAT = Law School Admission Test. These correlations are sample-size-weighted averages across the 21 schools. LSAT was adjusted for direct range restriction via procedures for Thorndike Case II; GRE was adjusted for indirect range restriction via procedures for Thorndike Case III.

Table 5 Sample-Size-Weighted Averages of the Validity of GRE and Law School Admission Test (LSAT) Over and Above the Validity of Undergraduate Grade Point Average (UGPA) With Adjusted *R* (and Adjusted R^2)

	Group I	Group II	Group III	All groups
Number of schools	6	8	7	21
Number of students	569	476	542	1,587
UGPA	.20 [.05]	.21 [.08]	.24 [.07]	.22 [.07]
+GRE-V, GRE-Q and GRE-AW separately entered	.41 [.17]	.39 [.18]	.45 [.21]	.42 [.19]
+GRE V + Q + AW composite (equally weighted)	.37 [.14]	.41 [.18]	.45 [.21]	.41 [.18]
+GRE V + Q composite (equally weighted)	.38 [.16]	.38 [.16]	.44 [.21]	.40 [.17]
+LSAT	.39 [.15]	.42 [.19]	.46 [.22]	.42 [.19]

Note. GRE-AW = GRE analytical writing section; GRE-Q = GRE quantitative reasoning section; GRE-V = GRE verbal reasoning section. In each of the rows containing a "+" symbol, UGPA and the scores to the right of the "+" symbol are included in the regression model. Ordinary least squares regression was used to determine validity. In each pair, the number on the left is the value of *R* (multiple correlation), and the number in square brackets to its right is an associated R^2 (coefficient of determination). The bracketed numbers are not squared values of the numbers to their left. Each school's *R* value was included in a sample-size-weighted averaging across the schools, and each school's corresponding R^2 value was included in a separate sample-size-weighted averaging across the schools. It is those separate averages that are reported in the table. These analyses do not adjust for restriction of range.

provided substantially more predictive information than UGPA did alone. For example, GRE-V, GRE-Q and GRE-AW separately entered into a regression increases the correlation for the prediction of LGPA by .12 over and above UGPA (i.e., $\Delta R^2 = .12 = .19$ for the GRE test minus .07 for UGPA). The ΔR^2 values for the GRE V + Q + AW composite and LSAT were similar (.18 and .19, respectively).

Schmidt and Hunter Meta-Analytic Evidence of Validity Generalization

Our final correlation-based analyses addressed the question of how generalizable the findings are for our 21 law schools by using the Schmidt and Hunter (2015) meta-analytic random effects approach, as described previously. In Table 6, $\hat{\bar{\rho}}$

Table 6 Schmidt and Hunter Validity Generalization Meta-Analyses

							80% CV		95% CI	
Predictor	N	k	$\bar{r}$	SD_r	$\hat{\bar{\rho}}$	SD_ρ	LL	UL	LL	UL
GRE V + Q + AW	1,587	21	.35	.10	.53[a]	.00	.53	.53	.47	.59
GRE V + Q	1,587	21	.33	.12	.54[a]	.00	.54	.54	.48	.60
LSAT	1,587	21	.33	.12	.55	.09	.44	.66	.48	.63

Note. N = sample size; k = number of schools; $\bar{r}$ = sample-size-weighted mean correlation; SD_r = sample-size-weighted standard deviation of observed correlations; $\hat{\bar{\rho}}$ = estimate of the mean true score correlation equal to the sample-size-weighted mean of the correlations for the 21 schools, with each correlation adjusted to account for range restriction; SD_ρ = true score standard deviation; CV = credibility interval; CI = confidence interval; LL = lower bounds; UL = upper bounds; CV_{LL} and CV_{UL} = lower and upper bounds, respectively, of the 80% CV; CI_{LL} and CI_{UL} = lower and upper bounds, respectively, of the 95% CI. As in prior analyses, Thorndike Cases II and III were used to adjust for range restriction. Sample-size-weighting was used in these meta-analyses.
[a]It is possible that the addition of a score (GRE-AW) to a composite (GRE V + Q) will reduce prediction of an outcome (LGPA) even if use of that score alone is predictive of that outcome. That could occur if the amount of statistical adjustment to account for sampling error and/or range restriction for the newly added score is less than the adjustment for the score(s) that already were in the composite. In the case of GRE-AW, the adjustment to account for range restriction was notably smaller than it was in the case of GRE-V or GRE-Q. The adjustment for GRE-AW was smaller, because the Thorndike Case III adjustment for indirect range restriction necessarily depends on the size of the correlation between the GRE section(s) in question and the LSAT. Since the LSAT lacks a scored writing section, GRE-AW was the GRE section with the weakest relationship to the LSAT, and thus it received the smallest adjustment to account for range restriction.

is the estimate of the mean true score correlation after accounting for range restriction. For calculating $\hat{\bar{\rho}}$, we made the adjustments by

1. using the appropriate Thorndike (1949) formula (Case III to account for GRE indirect range restriction and Case II to account for LSAT direct range restriction) to adjust each school's correlation between test scores (GRE or LSAT) and LGPA, and then
2. taking a sample-size-weighted average across the correlations after we adjusted them to account for range restriction.

The values of $\hat{\bar{\rho}}$ for the GRE (.53 and .54) reiterate the findings of strong evidence of validity reported in Table 4 for all groups after accounting for range restriction. Following the Schmidt and Hunter (2015) methodology, in Table 6 we report 95% confidence intervals for the GRE V + Q + AW (.47, .59), GRE V + Q (.48, .60), and the LSAT (.48, .63) correlation coefficients that are virtually identical to the confidence intervals that we calculated using bootstrapped methods reported previously (for the GRE V+ Q + AW correlation coefficient [.47, .57], and for the LSAT correlation coefficient [.47, .60]). As explained further in Appendix B, confidence intervals give a range of likely values for the true average range-restricted-adjusted correlation coefficient for the population of law schools given the uncertainty in the estimated value, because the estimation relies on a sample of law schools and students from those schools. When the confidence intervals do not include small values, then there is greater evidence that validity coefficients are on average not small. The key new findings in Table 6 are for the 80% credibility interval, which accounts for range restriction as well as sampling error. The 80% credibility interval is a range of values surrounding $\hat{\bar{\rho}}$, from 1.28 SD_ρ below $\hat{\bar{\rho}}$ (the lower limit) to 1.28 SD_ρ above $\hat{\bar{\rho}}$ (the upper limit). The value 1.28 is a multiplier that represents the 90th percentile for a standard normal probability distribution. Consequently, roughly 80% of the distribution of true score correlations will be contained in the interval. The true score correlation is the correlation for the entire population of applicants to each law school. We cannot observe those correlations, but by using the estimates from the sampled schools we can estimate the variance of their distribution using SD_ρ, the true score standard deviation. SD_ρ equals the square root of

1. the variance of the validity coefficients adjusted to account for range restriction, minus
2. the variance of these validity coefficients due to sampling error.[10]

The 80% CV columns in Table 6 indicate that the value of $\hat{\bar{\rho}}$ for each predictor is extremely stable (i.e., generalizes across the schools), because the difference between the 80% CV lower and upper limits is very small, essentially zero. Appendix B provides some additional detail about Schmidt and Hunter (2015) meta-analysis,

Table 7 Comparison of Top Third Versus Bottom Third of First-Year Law School Grades (LGPA) Based on Top Third Versus Bottom Third of Performance on the GRE V + Q + AW Composite (Equally Weighted)

	Group I	Group II	Group III	All groups
Number of schools	6	8	7	21
Number of students	569	476	542	1,587
Top third GRE and top third LGPA	45.0%	44.9%	46.7%	45.6%
Top third GRE and bottom third LGPA	22.2%	24.1%	18.9%	21.6%
Bottom third GRE and bottom third LGPA	46.8%	51.3%	48.9%	48.9%
Bottom third GRE and top third LGPA	24.5%	22.4%	20.6%	22.5%

particularly credibility intervals. In summary, the validity of the GRE test is strong and generalizes across law schools.

The validity coefficients in Tables 3–6 show that the GRE test is valid for predicting student performance in law school. This study's aggregate observed correlation coefficients in the .30s and adjusted correlation coefficients in the .50s for GRE V + Q + AW and V + Q compare favorably to the findings in past large-scale validity studies of postsecondary admission tests, including validity studies of the GRE and LSAT (cf. Anthony et al., 2016; Kuncel et al., 2001; Kuncel et al., 2010; Kuncel & Hezlett, 2007). Correlation coefficients in the middle .30s were in the top third in size among the effect sizes (correlation coefficients) reported in psychological assessment (see Hemphill, 2003) and in two of the leading journals on human performance assessment (see Bosco, Aguinis, Singh, Field, & Pierce, 2015); correlation coefficients were well into that top third (see Bosco et al., 2015; Hemphill, 2003). As demonstrated, even when one already possesses UGPA to make an admissions decision, the GRE test provides additional value.

Given the strong validity findings, one would expect the test to be very useful. In some statistical models of test utility, test validity interacts with other determinants of test usefulness (e.g., institutional selectivity and the number of selectees) to determine the test's utility (Klieger et al., 2014). The Taylor and Russell (1939) utility model indicates that the proportion of individuals who are selected based on test scores and later deemed satisfactory performers in future schooling is greatest when the selection ratio (admission rate) is low (and the test validity is high). Many law schools are selective, with hundreds of students selected each year and high graduation rates (see ABA, 2017c). The Hunter and Hunter (1984) utility model indicates that use of a test can provide increasing financial gains to an institution when there are increases in any of the following determinants (in addition to test validity): selectee tenure (the length of time students stay in school), the number of selectees, or the economic value that selectees provide to the institution. These determinants of utility are multiplicative such that utility gains from high levels of these determinants are even larger with higher levels of selection test validity.

Contingency Table Evidence of Validity

However, it is not always easy to understand the meaning and value of a particular correlation coefficient, and the utility from using a test is not always clear from the size of a correlation coefficient (see Klieger et al., 2014). Correlation and regression coefficients have little intrinsic meaning. Squaring a coefficient of .40 to find 16% of the variance accounted for might suggest that the predictor is not very useful as it does not account for 84% of the variance. Bridgeman et al. (2009) used contingency tables to show that correlation coefficients of this size and even smaller indeed have substantial practical value. In this approach, the success (in terms of high or low grades) of students with relatively high GRE scores is compared to the success of students with relatively low GRE scores. Specifically, in each school we identified students with GRE composite scores in the top and bottom third of the score distribution at that school, then we identified how many top third students (in terms of GRE scores) were in the top versus the bottom third of the LGPA distribution at that school. We did the same for students in the bottom third of the GRE scores at that school. The weighted average of these percentages, computed across the 21 schools, are presented in Table 7. For students in the top GRE third, about twice as many were in the top third of their law school classes as in the bottom GRE third. Similarly, for students in the bottom third of their classes in terms of GRE scores, about twice as many of these students were in bottom third in terms of LGPA as in the top GRE third.

Hierarchical Linear Modeling Evidence of Validity

When studies employ regression-based statistical methods to show predictive validity (with R multiple correlations or R^2 coefficients of determination), they typically use ordinary least squares (OLS) analyses, which assume independence of the residuals (Cohen, Cohen, West, & Aiken, 2003). OLS methods generally are appropriate when determining predictive validity for an individual school, but they can be less optimal if students from all schools are treated as if they were in just one group. In other words, law students who attend a particular law school are likely to be more similar to each other in ways relevant to predicting LGPA (e.g., more similar in skill levels, motivations, the distribution [e.g., grading curve] on which they are assessed relevant to predicting law school academic performance) than they are to attendees of another randomly chosen law school.

We used HLM, sometimes also referred to as multi-level modeling (MLM; Raudenbush & Bryk, 2002) to model possible effects of institutional differences in selectivity. We include the technical details of the HLM regression analyses in Appendix C. HLM methodology does not adjust for range restriction. The conclusions from these analyses confirm that the GRE test predicts LGPA whether one examines the GRE's usefulness alone, or its incremental value over and above the validity of using UGPA. The HLM analyses indicate that while the selectivity of the institution using the GRE test (or LSAT) has a statistically significant interaction effect on the size of the validity coefficient, the sizes of these effects are not practically significant. This finding is generally consistent with the conclusion from the Schmidt and Hunter (2015) meta-analyses that validity across schools is stable.

Conclusion

Our goals for this study were (a) to show the extent to which use of the GRE in law school admissions would satisfy the requirement of ABA Standard 503 that a law school use "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, 2014, p. 33) and (b) to abide by the best practice recommendations of the professional validation standards in educational testing (AERA et al., 2014; ETS, 2014a). The results of this national study showed that the GRE test scores have high reliability (Table 3). It was not possible to calculate GRE reliability specifically for the participants in the study, so we calculated it for the U.S. domestic GRE-taking population. We believe that this is a reasonable approach, because the GRE allows test takers to consider multiple forms of advanced education that include law school, business school, and many types of graduate school education. We could conclude that the level of reliability specifically for the participants was sufficient, because as we have observed previously, reliability sets the ceiling for criterion-related validity. Validity, the extent to which the use and interpretation of GRE scores are appropriate for predicting law school performance, is the key issue, and the levels of criterion-related validity that we found were substantial.

Given that validity is the central psychometric concern for reasons explained above, the primary research question in this study is the following: To what extent are GRE-V, GRE-Q, and GRE-AW scores valid and reliable measures of law school applicants' capabilities? For this study, we used LGPA as evidence of applicants' capabilities. We examined the relationship of test scores to law school grades for each GRE section alone, when scores for the different GRE sections were combined, and when GRE section scores were considered in conjunction with, and over and above, UGPA. To determine the validity of using the GRE test for law school admissions, ETS worked with 21 U.S. law schools that vary in selectivity, geographic location, and size, as well both public and private status (see Tables 1 and 2; Figure 1). We examined validity based on these schools overall as well as for schools based on their selectivity, because past empirical research has shown that validity for a standardized admissions test can vary based on school selectivity.

The results in this study indicate that the GRE test predicts LGPA when used alone (see Tables 3 and 4), in conjunction with UGPA (see Table 5), or over and above the validity of UGPA (when comparing the validity for UGPA versus validity for UGPA and GRE combined; see Table 5). Analyses demonstrated GRE validity within and across law school selectivity levels (cf. Groups I, II, and III in Tables 4 and 5). Meta-analysis showed no variance in GRE mean true score correlations and thus that GRE validity findings generalize across law schools (see Table 6). Findings from non-correlational validation methods that we used for the contingency tables (Table 7) provided further support for the practical significance of the correlational findings. HLM regression analyses (see Appendix C), which tests for variance in regression slopes while statistically accounting for the nesting of students within law schools, indicate that the GRE test is valid for use in law school admissions and that there is no practically significant variation in the validity coefficients across different levels of school selectivity or across law schools.

We did not observe practically meaningful differences in validity or reliability estimates between the GRE and LSAT. The correlations between the GRE (both the GRE composites and individual sections) and the LSAT for the 1,587 students in this study are quite high, reaching a level of .85 between GRE V + Q and the LSAT (see Table 3).

The results of this study provide strong support for the value of using GRE scores to inform admissions into law schools. Nonetheless, not all issues could be addressed in one study. We suggest some areas worth exploring in future research. First, although our sample was geographically diverse and included public and private law schools across a broad range of selectivity levels, it was limited to 21 out of the more than 200 accredited law schools in the United States. On the other hand, law schools' curricula, especially in the first year, tend to involve the same course subjects and use the case method instructional approach (LSAC, 2018). Still, future validity studies should expand the sample to include as much of the law school population as reasonably can be obtained, with schools representing the fullest achievable range of schools.

In addition, we had to use samples of students who had not taken the GRE expressly for the purpose of applying to law school. One possible concern about the relevance of these scores would be a possible lack of motivation to do well. One group of students in our analyses took the GRE after they had already been admitted to law school. All these students were compensated for taking the GRE test; an additional incentive for most of these students was a financial reward if they came close to matching their score on the LSAT. In addition, students were informed that the GRE scores received would be real and could be used to apply to a school in the future. We looked for, but did not find, any evidence that the test was not taken seriously (e.g., random responding to test questions). Indeed, the high correlations with LGPA that we observed would not be possible if many students were responding randomly. Nevertheless, a financial incentive or the possible use of GRE scores at some future date is not the same as a law school admissions incentive.

Another concern in interpreting scores in this group is that the students did not prepare for the test as they would for a high-stakes admissions test. This absence of preparation could be especially problematic for the GRE-Q scores, as an intensive review of some basic mathematics concepts covered on the test might be expected to have an impact on scores. On the other hand, most students presumably had prepared for reading passage–based questions before they took the LSAT, and such preparation should also prepare these students for the similar reading-passage based questions on the GRE-V. Thus, we strongly endorse future research with motivated students who had an opportunity to fully prepare for the test, but we do not anticipate substantially different results in that sample.

The other group of students in our analyses did take the GRE as a high-stakes test for admissions to competitive programs, albeit not law school programs. These students were highly motivated and had an opportunity to prepare for the test. We have increased confidence in our conclusions because both groups, with their different motivations and test preparation strategies, exhibited similar high correlations with LGPA. And when we examined LSAT–LGPA correlation coefficients from both groups, they were comparable to the correlations observed for the GRE. Even so, research on students who submitted GRE scores for the purpose of admissions to law school (when such scores become available) would provide additional evidence on the value of GRE scores.

A correlation between LGPA and scores collected for GRE tests taken postmatriculation could result from sources that would not exist for scores collected for GRE tests taken prior to matriculation, such as knowledge gained during law school. However, empirical evidence indicates this is unlikely to be the case. First, GRE scores are correlated with LGPA for the subsample in which scores were collected for GRE tests taken prior to matriculation. The correlation coefficients for the pre- and postmatriculation samples are similar, although the values are slightly larger for the postmatriculation subsample. This could be an indication of bias (in the technical statistical sense), but the correlation between the LSAT and LGPA also is larger for the GRE postmatriculation subsample than it is for the prematriculation subsample. Thus, empirical evidence indicates that the difference between the correlation coefficients of the two subsamples is likely related to the sample and not to the time at which the GRE test was taken.

If and when more law schools accept the GRE for admissions to law school and more students take the GRE for admission to law school, research that is closer to a population survey than a representative sample might provide additional insights. However, we would not expect the basic conclusions to change, as the current findings are very robust across law schools; the Schmidt and Hunter (2015) meta-analyses and the HLM analyses indicated remarkable consistency across the diverse set of law schools in our sample in the ability of GRE scores to predict LGPA.

Test validation is an ongoing process, and we look forward to what might be learned in future research. Nevertheless, the results in this study collectively provide strong empirical support for claims of reliability and predictive validity of the GRE for use in admissions at U.S. law schools.

Acknowledgments

The authors would like to thank the action editor, reviewers, and copyeditors for their helpful feedback. The authors are grateful as well to the law school deans, admission directors, and staff of the participating law schools for their efforts to secure and to provide the data that made this study possible.

Notes

1. With analogous considerations in mind, business schools worked with the GRE program to conduct similar research (see Young et al., 2014).
2. Reporting validity as correlations (r, R, or ρ) or coefficients of determination (R^2) is standard test validation practice, and it provides a quick way to compare different predictors and combinations of predictors.
3. Internal consistency is a way to measure the extent to which test questions that are designed to measure the same skills contribute to similar test scores. Item response theory is a psychometric approach to designing, analyzing, and scoring tests that is based on the relationship between performance on a test question and the skills that the test is intended to measure. MSTs are computer-based adaptive tests that select a group of test questions based on how well a test taker performed on the prior group of test questions.
4. R^2 validity estimates are calculated based on data from an available sample. Ultimately, one's interest is in validity for the population from which the sample comes. The R^2 for the sample may reflect idiosyncrasies of the sample that do not hold for the larger population. Therefore, R^2 based on the sample may overestimate R^2 for the population. Statistical adjustments can be made to R^2 to account for this overestimation. Unlike the unadjusted R^2, the adjusted R^2 value can decrease for each independent variable (e.g., an additional score, UGPA) that is added to a regression equation.
5. Range restriction is a restriction in the range of test scores (GRE, LSAT) and/or criterion values (LGPA), because admitted students reflect a narrower range of scores than would be found in the applicant pool. Ultimately, one wishes to know validity for law schools' applicant pools and not just for the students who attended. It also is possible that there is an effect on validity coefficients because of accepted students deciding not to attend.
6. Sackett and Yang (2000) described the assumptions underlying range restriction adjustments (a.k.a. corrections) for Thorndike (1949) Case II as follows: The correction formula requires two assumptions: that the x-y relationship is linear throughout the range of scores (i.e., the assumption of linearity) and that the error term is the same in the restricted sample and in the population (i.e., the assumption of homoscedasticity). Note that no normality assumption is required for the formula (Lawley, 1943). Note also that this formulation assumes nothing about the manner in which the selection occurs. Either or both tails of the distribution can be truncated, or only individuals in the tails of the distribution may be selected. (p. 114) Sackett and Yang indicated that Case III makes the same linearity and homoscedasticity assumptions as Case II does (p. 115). Research by Coward and Sackett (1990) and Cullen, Hardison, and Sackett (2004) indicates that cognitive ability – performance relationships typically are linear. Greener and Osburn (1979) determined that corrections for range restriction are robust when the homoscedasticity assumption does not hold (but that the corrections are not robust to more than small violations of the linearity assumption).
7. Specifically, we used 25,000 bootstrap samples where each was obtained by randomly sampling candidates from each institution with replacement. The 95% confidence interval for the correlation was obtained by the 0.025 and 0.975 quantiles of the distribution of the adjusted correlations across bootstrap samples.
8. An observed correlation is based directly on the original data, and it has not been statistically adjusted. A correlation (r) indicates the strength of the relationship of one variable (e.g., GRE-V) to another variable (e.g., UGPA), as compared to a multiple correlation (R) or coefficient of determination (R^2) that indicates the relationship of multiple variables together (e.g., GRE-V and GRE-Q together) to another variable (e.g., UGPA).
9. Note that "population" used here refers to the full set of institutions included in our study; it does not refer to the population of every law school in the United States.
10. The average unadjusted correlation across schools ($\bar{r}$) is used to calculate sampling error. The calculation involves three main steps. First, the sampling error estimation formula $\left(1 - \bar{r}^2\right)^2 / \left(N_i - 1\right)$ is used for each school, i. Since range restriction affects the estimation of sampling error, a second step is to refine the estimate of sampling error for each school based on the magnitude of range restriction affecting each school's unadjusted validity coefficient (for details, see Schmidt & Hunter, 2015). As the third step, a sample-size-weighted average of the refined estimates across schools can then be calculated.

11 Making range restriction adjustments is a step in conducting Schmidt & Hunter meta-analysis. Given that we possessed the data needed to use the Thorndike (1949) methods that Schmidt and Hunter (2015) consider optimal, we employed the Thorndike (1949) methods when we conducted the Schmidt and Hunter (2015) meta-analytic procedures.

12 It should be noted that the covariance matrices of random effects in Models 2a through 6d were restricted to be diagonal matrices; i.e. the variances of the random components are estimated but not the covariances.

References

Allen, J., & Robbins, S. (2010). Effects of interest-major congruence, motivation, and academic performance on timely degree attainment. *Journal of Counseling Psychology, 57*(1), 23–35. https://doi.org/10.1037/a0017267

American Bar Association. (2014). *2014–2015 ABA standards and rules of procedure for approval of law schools*. Chicago, IL: Author.

American Bar Association. (2015, January). *Managing director's guidance memo. Standard 503 and interpretation 503-3*. Retrieved from https://www.americanbar.org/content/dam/aba/administrative/legal_education_and_admissions_to_the_bar/governancedocuments/2015_s503_guidance_final.authcheckdam.pdf

American Bar Association. (2017a). *ABA-approved law schools*. Retrieved from https://www.americanbar.org/groups/legal_education/resources/aba_approved_law_schools.html

American Bar Association. (2017b). *Pre-law: Preparing for law school*. Retrieved from https://www.americanbar.org/groups/legal_education/resources/pre_law.html

American Bar Association. (2017c). *Section of legal education: ABA required disclosures. 2016*. Retrieved from http://www.abarequireddisclosures.org/Disclosure509.aspx

American Educational Research Association, American Psychological Association, & National Council on Measurement in Education. (2014). *Standards for educational and psychological testing*. Washington, DC: American Educational Research Association.

Anthony, L. C., Dalessandro, S. P., & Trierweiler, T. J. (2016, March). *Predictive validity of the LSAT: A national summary of the 2013 and 2014 LSAT correlation studies* (LSAT Technical Report 16-01). Newtown, PA: Law School Admission Council.

Bosco, F. A., Aguinis, H., Singh, K., Field, J. G., & Pierce, C. A. (2015). Correlational effect size benchmarks. *Journal of Applied Psychology, 100*, 431–449. https://doi.org/10.1037/a0038047

Bridgeman, B., Burton, N., & Cline, F. (2009). A note on presenting what predictive validity numbers mean. *Applied Measurement in Education, 22*, 109–119. https://doi.org/10.1080/08957340902754577

Briel, J., & Michel, R. (2014). Revisiting the GRE General Test. In C. Wendler & B. Bridgeman (Eds.), *The research foundation for the GRE revised General Test: A compendium of studies* (pp. 1.1.1–1.1.18). Retrieved from http://www.ets.org/s/research/pdf/gre_compendium.pdf

Cohen, J., Cohen, P., West, S. G., & Aiken, L. S. (2003). *Applied multiple regression/correlation analysis for the behavioral sciences* (3rd ed.). Mahwah, NJ: Routledge.

Coward, W. M., & Sackett, P. R. (1990). Linearity of ability-performance relationships: A reconfirmation. *Journal of Applied Psychology, 75*, 297–300. https://doi.org/10.1037/0021-9010.75.3.297

Cullen, M. J., Hardison, C. M., & Sackett, P. R. (2004). Using SAT-grade and ability-job performance relationships to test predictions derived from stereotype threat theory. *Journal of Applied Psychology, 89*, 220–230. https://doi.org/10.1037/0021-9010.89.2.220

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* (Technical Report No. 14-02). Newtown, PA: Law School Admission Council.

Educational Testing Service. (2014a). *ETS standards for quality and fairness*. Princeton, NJ: Author.

Educational Testing Service. (2014b). *A snapshot of the individuals who took the GRE revised General Test: August 2011–June 2014*. Retrieved from https://www.ets.org/s/gre/pdf/snapshot_test_taker_data_2014.pdf

Efron, B. (1979). Bootstrap methods: Another look at the jackknife. *The Annals of Statistics, 7*, 1–26. https://doi.org/10.1214/aos/1176344552

Enders, C. K., & Tofighi, D. (2007). Centering predictor variables in cross-sectional multilevel models: A new look at an old issue. *Psychological Methods, 12*, 121–138. https://doi.org/10.1037/1082-989X.12.2.121

Furr, R. M., & Bacharach, V. R. (2008). *Psychometrics: An introduction*. Thousand Oaks, CA: Sage.

Gabaldon, T. A. (2014). Doing the numbers: The numerate lawyer and transactional law. *American University Business Law Review, 3,* 63–80.

Greener, J. M., & Osburn, H. G. (1979). An empirical study of the accuracy of corrections for restriction in range due to explicit selection. *Applied Psychological Measurement, 3*(1), 31–41. https://doi.org/10.1177/014662167900300104

Hemphill, J. F. (2003). Interpreting the magnitudes of correlation coefficients. *American Psychologist, 58,* 78–79. https://doi.org/10.1037/0003-066X.58.1.78

Hunter, J. E., & Hunter, R. F. (1984). Validity and utility of alternative predictors of job performance. *Psychological Bulletin, 96*, 72–98. https://doi.org/10.1037/0033-2909.96.1.72

In pilot program, Harvard Law will accept GRE for admission: Accepting the GRE is part of an HLS strategy to expand access to legal education. (2017, March 8). *Harvard Law Today*. Retrieved from https://today.law.harvard.edu/gre/

Klieger, D. M., Cline, F. A., Holtzman, S. L., Minsky, J. L., & Lorenz, F. (2014). *New perspectives on the validity of the GRE General Test for predicting graduate school grades* (GRE Board Research Report No. 14-03). Princeton, NJ: Educational Testing Service. https://doi.org/10.1002/ets2.12026

Kobrin, J. L., Patterson, B. F., Shaw, E. J., Mattern, K. D., & Barbuti, S. M. (2008). *The validity of the SAT for predicting first-year grade point average* (Research Report 2008-5). New York, NY: College Board.

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

Kuncel, N. R., Hezlett, S. A., & Ones, D. S. (2001). A comprehensive meta-analysis of the predictive validity of the Graduate Record Examinations®: Implications for graduate student selection and performance. *Psychological Bulletin, 127,* 162–181. https://doi.org/10.1037/0033-2909.127.1.162

Kuncel, N. R., Wee, S., Serafin, L., & Hezlett, S. A. (2010). The validity of the Graduate Record Examinations for master's and doctoral programs: A meta-analytic investigation. *Educational and Psychological Measurement, 70*, 340–352. https://doi.org/10.1177/0013164409344508

Law School Admission Council. (2017). *Law School Admission Test (LSAT): LSAT scores as predictors of law school performance*. Retrieved from https://www.lsac.org/jd/lsat/your-score/law-school-performance

Law School Admission Council. (2018). *Thinking about law school: Is there a standard law school curriculum?* Retrieved from https://www.lsac.org/jd/thinking-about-law-school/standard-curriculum

Lawley, D. N. (1943). A note on Karl Pearson's selection formulae. *Proceedings of the Royal Society of Edinburgh, 62* (Section A, Part I), 28–30.

Lee, V. E. (2000). Using hierarchical linear modeling to study social contexts: The case of school effects. *Educational Psychologist, 35*, 125–141. https://doi.org/10.1207/S15326985EP3502_6

Mattern, K. D., & Patterson, B. F. (2011a). *Validity of the SAT for predicting second-year grades: 2006 SAT validity sample* (Statistical Report 2011-1). New York, NY: College Board.

Mattern, K. D., & Patterson, B. F. (2011b). *Validity of the SAT for predicting third-year grades: 2006 SAT validity sample* (Statistical Report 2011-3). New York, NY: College Board.

Mattern, K. D., & Patterson, B. F. (2011c). *Validity of the SAT for predicting fourth-year grades: 2006 SAT validity sample* (Statistical Report 2011-7). New York, NY: College Board.

Murphy, K. R. (Ed.). (2003). *Validity generalization: A critical review*. Mahwah, NJ: Erlbaum.

Nunnally, J. C. (1978). *Psychometric theory* (2nd ed.). New York, NY: McGraw-Hill.

Powers, D. E. (2004). Validity of Graduate Record Examinations (GRE) General Test scores for admissions to colleges of veterinary medicine. *Journal of Applied Psychology, 89*, 208–219. https://doi.org/10.1037/0021-9010.89.2.208

Raudenbush, S. W., & Bryk, A. S. (2002). *Hierarchical linear models: Applications and data analysis methods* (2nd ed.). Thousand Oaks, CA: Sage.

van Rijn, P. (2016). Reliability of multistage tests using item response theory. In D. Yan, A. von Davier, & C. Lewis (Eds.), *Computerized multistage testing: Theory and applications* (pp. 251–263). Boca Raton, FL: Chapman and Hall/CRC.

Robin, F., & Kim, S. (2014). Evaluating and maintaining the psychometric and scoring characteristics of the revised analytical writing measure. In C. Wendler & B. Bridgeman (Eds.), *The research foundation for the GRE revised General Test: A compendium of studies* (pp. 2.3.1–2.3.10). Retrieved from http://www.ets.org/s/research/pdf/gre_compendium.pdf

Robin, F., & Steffen, M. (2014). Test design for the GRE. In C. Wendler & B. Bridgeman (Eds.), *The research foundation for the GRE revised General Test: A compendium of studies* (pp. 3.3.1–3.3.12). Retrieved from http://www.ets.org/s/research/pdf/gre_compendium.pdf

Sackett, P. R., & Yang, H. (2000). Correction for range restriction: An expanded typology. *Journal of Applied Psychology, 85*, 112–118. https://doi.org/10.1037/0021-9010.85.1.112

Schmidt, F. L., & Hunter, J. E. (2015). *Methods of meta-analysis: Correcting error and bias in research findings* (3rd ed.). Thousand Oaks, CA: Sage. https://doi.org/10.4135/9781483398105

Schmidt, F. L., Viswesvaran, C., Ones, D. S., & Le, H. (2017). A failed challenge to validity generalization: Addressing a fundamental misunderstanding of the nature of VG. *Industrial and Organizational Psychology, 10,* 488–495. https://doi.org/10.4135/9781483398105

Taylor, H. C., & Russell, J. T. (1939). The relationship of validity coefficients to the practical validity of tests in selection. *Journal of Applied Psychology, 23,* 565–578. https://doi.org/10.1037/h0057079

Thorndike, R. L. (1949). *Personnel selection: Test and measurement techniques*. New York, NY: Wiley.

Wecker, M. (2012, January 19). Law review leads to legal jobs, recruiters say. *U.S. News & World Report*. Retrieved from https://www.usnews.com/education/best-graduate-schools/top-law-schools/articles/2012/01/19/law-review-leads-to-legal-jobs-recruiters-say

Weiss, D. C. (2013, October 29). Posner: Lawyers bad at math are an increasing concern; inmate's blood-pressure suit shows why. *ABA Journal*. Retrieved from http://www.abajournal.com/news/article/posner_math_block_lawyers_an_increasing_concern_inmates_blood-pressure_suit/

Wendler, C., & Bridgeman, B. (Eds.) (with C. Ezzo). (2014). *The research foundation for the GRE revised General Test: A compendium of studies*. Retrieved from https://www.ets.org/s/research/pdf/gre_compendium.pdf

Wightman, L. F. (2000). *Beyond FYA: Analysis of the utility of LSAT scores and UGPA for predicting academic success in law school* (Research Report No. 99-05). Newtown, PA: Law School Admission Council.

Yan, D., von Davier, A. A., & Lewis, C. (Eds.). (2014). *Computerized multistage testing: Theory and applications*. Boca Raton, FL: CRC Press.

Young, J. W., Klieger, D., Bochenek, J., Li, C., & Cline, F. (2014). *The validity of scores from the GRE revised General Test for forecasting performance in business schools: Phase one* (GRE Board Research Report No. 14-01). Princeton, NJ: Educational Testing Service. https://doi.org/10.1002/ets2.12019

Appendix A

Acknowledgments

We wish to acknowledge (in alphabetical order) the participation of the following law schools in this national study:

Boston University School of Law
Brigham Young University J. Reuben Clark Law School
Columbia Law School
Duke University School of Law
Emory University School of Law
Gonzaga University School of Law
The John Marshall Law School
Northwestern University Pritzker School of Law
The Notre Dame Law School
St. John's University School of Law
Suffolk University Law School
Texas A&M University School of Law
Thomas Jefferson School of Law
The University of Arizona James E. Rogers College of Law
The University of California, Davis, School of Law
The University of California, Los Angeles, School of Law
The University of Hawaii William S. Richardson School of Law
The University of Southern California
The University of Virginia School of Law
Wake Forest University School of Law
Washington University in St. Louis School of Law

Appendix B

Schmidt and Hunter Meta-Analytic Random-Effects Approach

The Schmidt and Hunter (2015) meta-analytic random-effects approach is a statistical methodology that asks what the validity of the indicator is for the population of interest (i.e., what the observed validity would be if there were no error in our validity estimates). In this study, we accounted for error due to the size of the sample (sampling error) and restriction of range,[11] as these were the relevant error sources. These sources of error can be thought of as interfering with our ability to see the true relationship between GRE (and LSAT) scores and LGPA. This meta-analytic approach estimates the validity for the population as $\hat{\bar{\rho}}$ (the Greek letter "rho" with an overline (¯), or "bar" symbol, above it to indicate that it is a mean [average] and a circumflex (^), or "hat" symbol, above it to indicate that it is an estimate). Furthermore, this meta-analytic approach estimates the extent to which the observed validity coefficients would vary across schools, if the statistical artifacts did not obscure them. If variability is small, then one may infer that the validity of the test generalizes — hence the use of the term “validity generalization” to describe this methodology. Educational research organizations and validity researchers generally support the judicious use of validity generalization methods (see AERA et al., 2014; Murphy, 2003; Schmidt & Hunter, 2015).

The credibility interval is a way in which the Schmidt and Hunter (2015) method indicates whether validity generalizes. The credibility interval reflects the idea that the true validity coefficient for a test could be different across different schools. It traditionally is set to 80%, indicating that 80% of the values in the distribution of true validity are within the credibility interval (Schmidt & Hunter, 2015). The lower bound of the credibility interval indicates a “worst-case” situation for validity, because there is a 90% probability that validity is larger than that lower bound value (Schmidt, Viswesvaran, Ones, & Le, 2017). It is when the lower bound (LL) for the credibility interval contains 0 that validity generalization typically comes into question.

The credibility interval is not a confidence interval. A confidence interval (traditionally often set to 95%) estimates the amount of error in the estimated mean (average) value of true validity, $\hat{\bar{\rho}}$, due to sampling error. This interval surrounds a single value, $\hat{\bar{\rho}}$. A confidence interval tries to answer the question, given the sampling error in the estimate, of what range of values is 95% likely to contain the true average validity coefficient for law schools. Conversely, a credibility interval does not reflect sampling error, because sampling error is not a part of SD_ρ: In calculating the SD_ρ on which the credibility interval is based, an estimate of sampling error is subtracted from the variance of the validity coefficients adjusted to account for range restriction. Whereas a confidence interval surrounds a single value, $\hat{\bar{\rho}}$, a credibility interval contains the 80% of the distribution of ρ values. A credibility interval is a type of moderator analysis that tries to answer the question of whether validity for the data on which the credibility interval is based is distributed narrowly enough to support the conclusion that the validity coefficient $\hat{\bar{\rho}}$ applies to all of the law schools within the data as a single population of schools. If the credibility interval is not narrow enough and consequently that conclusion is not supportable, then the data should be split up into hypothesized subpopulations of law schools and then re-meta-analyzed separately for each of these subpopulations. For example, a credibility interval containing 0 might indicate that the validity of a test when used by less selective schools is very different than its validity when used by very selective schools. If a selectivity moderator is detected, then data for less selective and highly selective schools should be re-meta-analyzed separately to estimate validity for each selectivity subpopulation and whether other moderators are present within either subpopulation.

Appendix C

Hierarchical Linear Modeling/Multilevel Validity Analyses

Research Questions

1. What proportion of the variance in LGPA is accounted for by institutional variation?
2. What is the predictive validity of GRE when used alone (and, for comparison, LSAT when used alone)?
3. What is the added predictive validity of GRE (and for LSAT) over and above UGPA?
4. Is the relationship between GRE and LGPA moderated by institutional selectivity? (Is the relationship between LSAT and LGPA moderated by institutional selectivity?)

Method

Participants

The sample consisted of 1,587 students enrolled in 21 law schools. The average cluster size (number of students within a law school) was 76 students, ranging from 34 to 143.

Variables

Outcome (dependent) variable. In all models, LGPA was used as the outcome variable.

Level 1 (predictor) variables. The Level 1 variables are variables that describe individuals. Level 1 variables used were UGPA, GRE-V score, GRE-Q score, GRE-AW score, the unit-weighted composite score of the three GRE sections (GRE

V + Q + AW), the unit-weighted composite score of the verbal and quantitative GRE sections (GRE V + Q), and LSAT score.

Level 2 (predictor) variables. Level 2 variables describe the groups (i.e., schools) into which the individuals are clustered. Institutional selectivity was used as the Level 2 predictor. Selectivity was measured in two different ways across different exploratory models. First, a continuous variable was calculated containing the median institutional LSAT score for each law school (as reported in ABA, 2017c). Second, we used a categorical variable for membership in Groups I – III (as defined in the main text) to reflect differences in school selectivity.

Statistical Procedures

HLM models were estimated to answer the research questions. No data were missing. All analyses were conducted using the SAS software, with maximum likelihood estimation.

RQ1. What proportion of the variance in LGPA is accounted for by institutional variation? To answer this question we conducted a random effects ANOVA with no predictors (Model 1). The purpose of this model was to compute the intraclass correlation (ICC) for the outcome variable LGPA (i.e., the ratio of between-institution variance to total observed variance). The ICC represents the proportion of LGPA variance accounted for by institutional variability. ICC values above .10 indicate the need to use HLM to control for clustering (Lee, 2000). Following Raudenbush and Bryk (2002), Model 1 is expressed as in Equation C1:

Level 1 :

$$Y_{ij} = \beta_{0j} + r_{ij}$$

Level 2 :

$$\beta_{0j} = \gamma_{00} + \mu_{0j} \tag{C1}$$

Combined model :

$$Y_{ij} = \gamma_{00} + \mu_{0j} + r_{ij},$$

where Y_{ij} is LGPA for student *i* in institution *j*, γ_{00} is the LGPA grand mean, μ_{0j} is the deviation between institution *j*'s LGPA mean and the LGPA grand mean, and r_{ij} is the deviation between student *i*'s LGPA score and institution *j*'s LGPA mean. The model estimates variances for r_{ij} and μ_{0j}, labeled σ^2 (within-institution residual variance) and τ_{00} (between-institution variance), respectively.

RQ2. What is the predictive validity of GRE when used alone (and, for comparison, LSAT when used alone)? In order to analyze the predictive validity of GRE and the LSAT, they were each included as Level 1 variables in separate Models 2a through 2d. For models with GRE V + Q + AW, GRE V + Q, and LSAT as Level 1 predictors, the model is expressed as in Equation C2:

Level 1 :

$$Y_{ij} = \beta_{0j} + \beta_{1j}\left(\text{level 1 predictor}\right) + r_{ij}$$

Level 2 :

$$\begin{aligned}\beta_{0j} &= \gamma_{00} + \mu_{0j}\\ \beta_{1j} &= \gamma_{10} + \mu_{1j}\end{aligned} \tag{C2}$$

Combined model :

$$Y_{ij} = \gamma_{00} + \gamma_{10}\left(\text{level 1 predictor}\right) + \mu_{0j} + \mu_{1j}\left(\text{level 1 predictor}\right) + r_{ij},$$

where γ_{00} is the average of mean LGPA within institutions; γ_{10} is a fixed slope for GRE V + Q + AW, GRE V + Q, or LSAT across institutions; μ_{1j} is the deviation between institution *j*'s slope for GRE V + Q + AW, GRE V + Q, or LSAT and the fixed slope across institutions; and the remaining parameters are defined as in Equation C1. In addition to the estimation of σ^2 (within-institution residual variance) and τ_{00} (between-institution variance), the model estimates the variance of μ_{1j}[12], labeled τ_{11}; μ_{1j} indicates how much variability exists in the slope for GRE V + Q + AW, GRE V + Q, or LSAT across institutions.

In Model 2c (Equation C3, we examined the effect of the three GRE sections separately entered (rather than composited into a single variable):

Level 1 :

$$Y_{ij} = \beta_{0j} + \beta_{1j}\,(GREQ) + \beta_{2j}\,(GREV) + \beta_{3j}\,(GREAW) + r_{ij}$$

Level 2 :

$$\begin{aligned} \beta_{0j} &= \gamma_{00} + \mu_{0j} \\ \beta_{1j} &= \gamma_{10} + \mu_{1j} \\ \beta_{2j} &= \gamma_{20} + \mu_{2j} \\ \beta_{3j} &= \gamma_{30} + \mu_{3j} \end{aligned} \tag{C3}$$

Combined model :

$$\begin{aligned} Y_{ij} = {} & \gamma_{00} + \gamma_{10}\,(GREQ) + \gamma_{20}\,(GREV) + \gamma_{30}\,(GREAW) + \mu_{0j} + \mu_{1j}\,(GREQ) \\ & + \mu_{2j}\,(GREV) + \mu_{3j}\,(GREAW) + r_{ij}. \end{aligned}$$

In addition to the parameters explained in previous models, Model 2c estimates the variance of μ_{1j} labeled τ_{11} (variation in the slope of GREQ across institutions), the variance of μ_{2j} labeled τ_{22} (variation in the slope of GREV across institutions), and the variance of μ_{3j} labeled τ_{33} (variation in the slope of GREAW across institutions).

All Level 1 variables in Model 2 were centered within cluster (i.e., group-mean centering) following Enders and Tofighi (2007). The fit of Models 2a through 2d was compared to the fit of Model 1 (the unconditional or random effects ANOVA model) by looking at the difference in −2 Log Likelihood (deviance), which is distributed as a χ^2; a significant difference in deviance indicates that the model with more parameters (Models 2a–2d) fit the data significantly better than a model with fewer parameters (Model 1). The models were also evaluated by calculating the proportion of within-institution (Level 1) variance accounted for by the addition of Level 1 predictors in comparison to Model 1.

RQ3. What is the added predictive validity of GRE (and, for additional context, LSAT), over and above UGPA? In order to analyze the incremental predictive validity of the GRE sections and composites, and LSAT beyond that provided by UGPA, we first established the unique proportion of LGPA variance explained by UGPA. In Model 3a (Equation C4), UGPA was included as the only Level 1 predictor:

Level 1 :

$$Y_{ij} = \beta_{0j} + \beta_{1j}\,(UGPA) + r_{ij}$$

Level 2 :

$$\begin{aligned} \beta_{0j} &= \gamma_{00} + \mu_{0j} \\ \beta_{1j} &= \gamma_{10} + \mu_{1j} \end{aligned} \tag{C4}$$

Combined model :

$$Y_{ij} = \gamma_{00} + \gamma_{10}\,(UGPA) + \mu_{0j} + \mu_{1j}\,(UGPA) + r_{ij}.$$

GRE V + Q + AW and GRE V + Q were added as Level 1 predictors in Models 3b and 3c, respectively. LSAT was added as a Level 1 predictor in Model 3e. Models 3b, 3c, and 3e are expressed as in Equation C5:

Level 1 :

$$Y_{ij} = \beta_{0j} + \beta_{1j}\,(UGPA) + \beta_{2j}\left(level1\ predictor\right) + r_{ij}$$

Level 2 :

$$\begin{aligned} \beta_{0j} &= \gamma_{00} + \mu_{0j} \\ \beta_{1j} &= \gamma_{10} + \mu_{1j} \\ \beta_{2j} &= \gamma_{20} + \mu_{2j} \end{aligned} \tag{C5}$$

Combined model :

$$\begin{aligned} Y_{ij} = {} & \gamma_{00} + \gamma_{10}\,(UGPA) + \gamma_{20}\left(level1predictor\right) + \mu_{0j} + \mu_{1j}\,(UGPA) \\ & + \mu_{2j}\left(level1predictor\right) + r_{ij}. \end{aligned}$$

In Model 3d, all three GRE sections (GRE-Q, GRE-V, and GRE-AW) were included as separately entered Level 1 predictors. The model is expressed as in Equation C6:

$$\begin{aligned}
&\text{Level 1}: \\
&\quad Y_{ij} = \beta_{0j} + \beta_{1j}\,(UGPA) + \beta_{2j}\,(GREQ) + \beta_{3j}\,(GREV) + \beta_{4j}\,(GREAW) + r_{ij} \\
&\text{Level 2}: \\
&\quad \beta_{0j} = \gamma_{00} + \mu_{0j} \\
&\quad \beta_{1j} = \gamma_{10} + \mu_{1j} \\
&\quad \beta_{2j} = \gamma_{20} + \mu_{2j} \\
&\quad \beta_{3j} = \gamma_{30} + \mu_{3j} \\
&\quad \beta_{4j} = \gamma_{40} + \mu_{4j} \\
&\text{Combined model}: \\
&\quad Y_{ij} = \gamma_{00} + \gamma_{10}\,(UGPA) + \gamma_{20}\,(GREQ) + \gamma_{30}\,(GREV) + \gamma_{40}\,(GREAW) + \\
&\quad\quad \mu_{0j} + \mu_{1j}\,(UGPA) + \mu_{2j}\,(GREQ) + \mu_{3j}\,(GREV) + \mu_{4j}\,(GREAW) + r_{ij}.
\end{aligned} \tag{C6}$$

The fit of Models 3b through 3e was compared to the fit of Model 3a using the difference in deviance between models. Further, the unique contribution of these variables in Models 3b through 3e was examined by comparing the proportion of Level 1 variance explained by Models 3b through 3e to the same figure calculated for Model 3a.

RQ4. Is the relationship between GRE and LGPA moderated by institutional selectivity? (And, for additional context, is the relationship between LSAT and LGPA moderated by institutional selectivity?) Without controlling for institutional variability as we do here, we examined these questions previously in this manuscript. In the prior analyses, we had categorized schools into three groups (Groups I–III) based on what made conceptual sense, and we had then examined validity coefficients separately for each group (as well as overall). In these HLM analyses, we investigated selectivity in two ways: (a) We modeled selectivity by including a continuous variable that measured the LSAT median for each school's entering class (Models 4a–4d), and (b) we separately modeled selectivity using categorical membership in Groups I–III (Models 5a–6d).

In all models, we included selectivity (measured as institutional median LSAT or as membership to Groups I–III) as a Level 2 predictor, and we added a cross-level interaction between selectivity and our Level 1 predictors (GRE V + Q + AW, GRE V + Q, each separate GRE section, and LSAT). Given our interest in the cross-level interaction, Level 1 predictors were centered within cluster, and median institutional LSAT was grand-mean centered (Enders & Tofighi, 2007).

Models 4a and 4b with GRE V + Q + AW and GRE V + Q as Level 1 predictors, respectively, and Model 4d with LSAT as a Level 1 predictor, are expressed as in Equation C7:

$$\begin{aligned}
&\text{Level 1}: \\
&\quad Y_{ij} = \beta_{0j} + \beta_{1j}\,(UGPA) + \beta_{2j}\left(level1predictor\right) + r_{ij} \\
&\text{Level 2}: \\
&\quad \beta_{0j} = \gamma_{00} + \gamma_{01}\,(medianLSAT) + \mu_{0j} \\
&\quad \beta_{1j} = \gamma_{10} + \mu_{1j} \\
&\quad \beta_{2j} = \gamma_{20} + \gamma_{21}\,(medianLSAT) + \mu_{2j} \\
&\text{Combined model}: \\
&\quad Y_{ij} = \gamma_{00} + \gamma_{10}\,(UGPA) + \gamma_{20}\left(level1predictor\right) + \gamma_{01}\,(medianLSAT) + \\
&\quad\quad \gamma_{21}\left(level1predictor\right)(medianLSAT) + \mu_{0j} + \mu_{1j}\,(UGPA) + \\
&\quad\quad \mu_{2j}\left(level1predictor\right) + r_{ij}.
\end{aligned} \tag{C7}$$

Model 4c, which included an interaction between median institutional LSAT and each separate GRE section separately entered can be expressed as in Equation C8:

$$
\begin{aligned}
&\text{Level 1:}\\
&\quad Y_{ij} = \beta_{0j} + \beta_{1j}\,(UGPA) + \beta_{2j}\,(GREQ) + \beta_{3j}\,(GREV) + \beta_{4j}\,(GREAW) + r_{ij}\\
&\text{Level 2:}\\
&\quad \beta_{0j} = \gamma_{00} + \gamma_{01}\,(medianLSAT) + \mu_{0j}\\
&\quad \beta_{1j} = \gamma_{10} + \mu_{1j}\\
&\quad \beta_{2j} = \gamma_{20} + \gamma_{21}\,(medianLSAT) + \mu_{2j}\\
&\quad \beta_{3j} = \gamma_{30} + \gamma_{31}\,(medianLSAT) + \mu_{3j}\\
&\quad \beta_{4j} = \gamma_{40} + \gamma_{41}\,(medianLSAT) + \mu_{4j}\\
&\text{Combined model:}\\
&\quad Y_{ij} = \gamma_{00} + \gamma_{01}\,(medianLSAT) + \gamma_{10}\,(UGPA) + \gamma_{20}\,(GREQ) +\\
&\quad \gamma_{30}\,(GREV) + \gamma_{40}\,(GREAW) + \gamma_{21}\,(GREQ)\,(medianLSAT) +\\
&\quad \gamma_{31}\,(GREV)\,(medianLSAT) + \gamma_{41}\,(GREAW)\,(medianLSAT) + \mu_{0j}\\
&\quad +\mu_{1j}\,(UGPA) + \mu_{2j}\,(GREQ) + \mu_{3j}\,(GREV) + \mu_{4j}\,(GREAW) + r_{ij}.
\end{aligned}
\tag{C8}
$$

In addition to looking at selectivity as a continuous variable, we were also interested in whether selectivity acted as a moderator when groups of institutions were categorized as exhibiting low versus moderate versus high selectivity. As mentioned, we previously used this categorical approach to compare validity coefficients across school selectivity in a conceptually straightforward way. In Models 5a through 5d, selectivity was dummy coded using high selectivity as the reference group, such that Dummy 1 represented a comparison between high and moderate selectivity, and Dummy 2 represented a comparison between high and low selectivity. In order to also generate a comparison between low and moderately selective institutions, Models 6a through 6d specified low selectivity as the reference group. In these models, Dummy 1 represented a comparison between low and moderate selectivity, and Dummy 2 represented the same comparison as specified above (high vs. low selectivity).

Models 5a through 5b and 6a and 6b for GRE V + Q + AW and GRE V + Q, respectively, and Models 5d and 6d for LSAT are expressed as in Equation C9:

$$
\begin{aligned}
&\text{Level 1:}\\
&\quad Y_{ij} = \beta_{0j} + \beta_{1j}\,(UGPA) + \beta_{2j}\left(level1predictor\right) + r_{ij}\\
&\text{Level 2:}\\
&\quad \beta_{0j} = \gamma_{00} + \gamma_{01}\left(Dummy1\right) + \gamma_{02}\left(Dummy2\right) + \mu_{0j}\\
&\quad \beta_{1j} = \gamma_{10} + \mu_{1j}\\
&\quad \beta_{2j} = \gamma_{20} + \gamma_{21}\left(Dummy1\right) + \gamma_{22}\left(Dummy2\right) + \mu_{2j}\\
&\text{Combined model:}\\
&\quad Y_{ij} = \gamma_{00} + \gamma_{10}\,(UGPA) + \gamma_{20}\left(level1predictor\right) + \gamma_{01}\left(Dummy1\right) +\\
&\quad \gamma_{02}\left(Dummy2\right) + \gamma_{21}\left(level1predictor\right)\left(Dummy1\right) +\\
&\quad \gamma_{22}\left(level1predictor\right)\left(Dummy2\right) + \mu_{0j} + \mu_{1j}\,(UGPA) +\\
&\quad \mu_{2j}\left(level1predictor\right) + r_{ij}.
\end{aligned}
\tag{C9}
$$

In Models 5c and 6c, the interaction between each GRE section separately entered and dummy coded institutional selectivity was explored. These models are expressed as in Equation C10:

Level 1：

$$Y_{ij} = \beta_{0j} + \beta_{1j}(UGPA) + \beta_{2j}(GREQ) + \beta_{3j}(GREV) + \beta_{4j}(GREAW) + r_{ij}$$

Level 2：

$$\begin{aligned}
\beta_{0j} &= \gamma_{00} + \gamma_{01}\left(Dummy1\right) + \gamma_{02}\left(Dummy2\right) + \mu_{0j}\\
\beta_{1j} &= \gamma_{10} + \mu_{1j}\\
\beta_{2j} &= \gamma_{20} + \gamma_{21}\left(Dummy1\right) + \gamma_{22}\left(Dummy2\right) + \mu_{2j}\\
\beta_{3j} &= \gamma_{30} + \gamma_{31}\left(Dummy1\right) + \gamma_{32}\left(Dummy2\right) + \mu_{3j}\\
\beta_{4j} &= \gamma_{40} + \gamma_{41}\left(Dummy1\right) + \gamma_{42}\left(Dummy2\right) + \mu_{4j}
\end{aligned} \tag{C10}$$

Combined model：

$$\begin{aligned}
Y_{ij} = {} & \gamma_{00} + \gamma_{01}\left(Dummy1\right) + \gamma_{02}\left(Dummy2\right) + \gamma_{10}(UGPA) + \\
& \gamma_{20}(GREQ) + \gamma_{30}(GREV) + \gamma_{40}(GREAW) + \\
& \gamma_{21}(GREQ)\left(Dummy1\right) + \gamma_{22}(GREQ)\left(Dummy2\right) + \\
& \gamma_{31}(GREV)\left(Dummy1\right) + \gamma_{32}(GREV)\left(Dummy2\right) + \\
& \gamma_{41}(GREAW)\left(Dummy1\right) + \gamma_{42}(GREAW)\left(Dummy2\right) + \mu_{0j} + \\
& \mu_{1j}(UGPA) + \mu_{2j}(GREQ) + \mu_{3j}(GREV) + \mu_{4j}(GREAW) + r_{ij}.
\end{aligned}$$

In total, we conducted 22 HLM models. See Table C1 for a summary of the models examined.

Results

*RQ1. What proportion of the variance in LGPA across the participating students is accounted for by institutional variation (*i.e.*, the fact that students attended different law schools)?*

The ICC for the random effects ANOVA revealed that 27% of the overall variance in LGPA was attributable to institutional differences (Table C2). HLM was therefore implemented in subsequent analyses to partition variability at the student versus institutional levels.

RQ2. What is the independent predictive validity of GRE (and, for additional context, LSAT)?

Model 2c resulted in a nonpositive definite matrix due to the estimation of τ_{22}, indicating that there was no variability in the slope of GRE-V across institutions; hence, τ_{22} was fixed to zero, and the model was re-estimated. The fit of Models 2a through 2d is reported in Table C2. According to the difference in deviance, Models 2a through 2d showed a better fit to the data than Model 1. The main effects for GRE V + Q + AW, GRE V + Q, all three GRE sections entered separately into the same HLM model, and LSAT were statistically significant. In comparison to Model 1, the addition of these variables explained 13.02%, 12.43%, 13.90%, and 12.88% of Level 1 variance in LGPA in Models 2a through 2d, respectively. Overall, these results show the independent predictive validity of the GRE and LSAT in the prediction of LGPA.

Models 2b, 2c, and 2d show statistically significant variation in slopes on GRE scores across law schools. Results from Schmidt and Hunter (2015) analyses showed small or no variance in the correlations as reflected in the credibility intervals with close or equal upper and lower bounds. There are some differences between the HLM and Schmidt and Hunter methods: The Schmidt and Hunter results use school-specific correlations that were adjusted for restriction in range. The HLM model assumes constant variance in LGPA. If that is true and the correlations are equal, then the variation in the slopes will result from variance in the reciprocal of the standard deviation in GRE scores within each school. The variance in the GRE scores is likely to vary across law schools, offering a possible explanation for differences in results.

RQ3. What is the added predictive validity of GRE (and, for additional context, LSAT) over and above UGPA?

Model 3d resulted in a nonpositive definite matrix due to the estimates of τ_{33} and τ_{44}, indicating that there was no variability in the slope of GRE-V and in the slope GRE-AW across institutions; hence, τ_{33} and τ_{44} were fixed to zero.

All models show a significantly better fit to the data than the model that only included UGPA (Model 3a), according to the difference in deviance (Table C3). Further, the results show that all Level 1 predictors were statistically significant.

Table C1 Summary of Hierarchical Linear Modeling Examined

Model	Level 1 predictors	Level 2 predictors	Random effects
1	—	—	Random intercepts for schools
2a	GREQVAW	—	Random intercepts + random slope for GREQVAW
2b	GREQV	—	Random intercepts + random slope for GREQV
2c	GREQ + GREV+ GREAW	—	Random intercepts + random slope for GREQ, GREV and GREAW
2d	LSAT	—	Random intercepts + random slope for LSAT
3a	UGPA	—	Random intercepts + random slope for UGPA
3b	UGPA + GREQVAW	—	Random intercepts + random slope for UGPA and GREQVAW
3c	UGPA + GREQV	—	Random intercepts + random slope for UGPA and GREQV
3d	UGPA + GREQ + GREV+ GREAW	—	Random intercepts + random slope for UGPA, GREQ, GREV and GREAW
3e	UGPA + LSAT	—	Random intercepts + random slope for UGPA and LSAT
4a	UGPA + GREQVAW	Median LSAT	Random intercepts + random slope for UGPA and GREQVAW
4b	UGPA + GREQV	Median LSAT	Random intercepts + random slope for UGPA and GREQV
4c	UGPA + GREQ + GREV+ GREAW	Median LSAT	Random intercepts + random slope for UGPA, GREQ, GREV and GREAW
4d	UGPA + LSAT	Median LSAT	Random intercepts + random slope for UGPA and LSAT
5a	UGPA + GREQVAW	Dummy1(high vs. medium selectivity) + Dummy2(high vs. low selectivity)	Random intercepts + random slope for UGPA and GREQVAW
5b	UGPA + GREQV	Dummy1(high vs. medium selectivity) + Dummy2(high vs. low selectivity)	Random intercepts + random slope for UGPA and GREQV
5c	UGPA + GREQ + GREV+ GREAW	Dummy1(high vs. medium selectivity) + Dummy2(high vs. low selectivity)	Random intercepts + random slope for UGPA, GREQ, GREV and GREAW
5d	UGPA + LSAT	Dummy1(high vs. medium selectivity) + Dummy2(high vs. low selectivity)	Random intercepts + random slope for UGPA and LSAT
6a	UGPA + GREQVAW	Dummy1(low vs. medium selectivity) + Dummy2(low vs. high selectivity)	Random intercepts + random slope for UGPA and GREQVAW
6b	UGPA + GREQV	Dummy1(low vs. medium selectivity) + Dummy2(low vs. high selectivity)	Random intercepts + random slope for UGPA and GREQV
6c	UGPA + GREQ + GREV+ GREAW	Dummy1(low vs. medium selectivity) + Dummy2(low vs. high selectivity)	Random intercepts + random slope for UGPA, GREQ, GREV and GREAW
6d	UGPA + LSAT	Dummy1(low vs. medium selectivity) + Dummy2(low vs. high selectivity)	Random intercepts + random slope for UGPA and LSAT

Note. GREAW = GRE analytical writing section; GREQ = GRE quantitative reasoning section; GREV = GRE verbal reasoning section; GREQV = GRE quantitative reasoning and verbal reasoning composite; GREQVAW = GRE quantitative reasoning, verbal reasoning, and analytical writing composite; LSAT = Law School Admission Test; UGPA = undergraduate grade point average.

Table C2 Hierarchical Linear Modeling Results With Level 1 Predictors: Predictive Validity of GRE (and LSAT) Scores

	Model				
	1	2a	2b	2c	2d
Fixed effects					
γ_{00}	3.22 (0.05)***	3.22 (0.05)***	3.22 (0.05)***	3.22 (0.05)***	3.22 (0.05)***
γ_{10}		0.07 (0.01)***	0.09 (0.01)***	0.01 (0.00)***	0.02 (0.00)***
γ_{20}				0.01 (0.00)***	
γ_{30}				0.06 (0.01)***	
Variance components					
τ_{00}	0.05 (0.02)**	0.05 (0.02)**	0.05 (0.02)**	0.05 (0.02)**	0.05 (0.02)**
τ_{11}		0.0003 (0.00)	0.001 (0.00)*	0.0001 (0.00) *	0.0001 (0.00)*
τ_{22}				--	
τ_{33}				0.0002 (0.00)	
σ^2	0.14 (0.00)***	0.12 (0.00)***	0.12 (0.00)***	0.12 (0.00)***	0.12 (0.00)***
Reduction in within level variance					
vs. Model 1		13.02%	12.43%	13.90%	12.88%
Fit of the model					
-2LogLikelihood	1,410.3	1,202.7	1,218.4	1,193.7	1,213.3
Δ Deviance (vs Model1)		207.6(2)***	191.9(2)***	216.6(5)***	197(2)***

Note. Standard errors are shown in parenthesis. Model 1: Random effects ANOVA. Parameters per model: Model 2a: γ_{10}=GRE V + Q + AW; Model 2b: γ_{10}=GRE V + Q; Model 2c: γ_{10}=GRE-Q, γ_{20}= GRE-V, γ_{30}= GRE-AW; Model 2d: γ_{10}=LSAT. The intraclass correlation (ICC) was computed as $\tau_{00}/(\tau_{00}+\sigma^2)$. In Model 1, $ICC = \frac{.05}{.05+.14} \approx .27$

The ICC and the reduction in within level variance was computed based on the nonrounded values of τ_{00} and σ^2. Hence, although the values reported of τ_{00} and σ^2 across models are identical in the table, the reduction in variance was different.

$^{***}p < .001$, $^{**}p < .01$, $^{*}p < .05$

Table C3 Hierarchical Linear Modeling Results With Level 1 Predictors: Predictive Validity of GRE (and LSAT) Over and Above UGPA

	Model				
	3a	3b	3c	3d	3e
Fixed effects					
γ_{00}	3.22 (0.05)***	3.22 (0.05)***	3.22 (0.05)***	3.22 (0.05)***	3.22 (0.05)***
γ_{10}	0.22 (0.04)***	0.20 (0.04)***	0.22 (0.04)***	0.21 (0.04)***	0.24 (0.04)***
γ_{20}		0.07 (0.00)***	0.09 (0.01)***	0.01 (0.00)***	0.02 (0.00)***
γ_{30}				0.01 (0.00)***	
γ_{40}				0.06 (0.01)***	
Variance components					
τ_{00}	0.05 (0.02)**	0.05 (0.02)**	0.05 (0.02)**	0.05 (0.02)**	0.05 (0.02)**
τ_{11}	0.02 (0.01)	0.02 (0.01)	0.02 (0.01)	0.03 (0.02)	0.02 (0.01)*
τ_{22}		0.0003 (0.00)	0.001 (0.00) *	0.0001 (0.00) *	0.0001 (0.00)*
τ_{33}				—	
τ_{44}				—	
σ^2	0.13 (0.00)***	0.11 (0.00)***	0.11 (0.00)***	0.11 (0.00)***	0.11 (0.00)***
Reduction in within level variance					
vs. Model 1	6.78%	18.97%	18.97%	20.36%	18.97%
vs. Model 3a		13.07%	13.07%	14.57%	13.07%
Fit of the model					
-2LogLikelihood	1,318.1	1,112.7	1,117.9	1,096.6	1,096.6
Δ Deviance (vs Model3a)		205.4(2)***	200.2(2)***	221.5(4)***	221.5(2)***

Note. Standard errors are shown in parenthesis. Parameters per model: Model 3a: γ_{10}=UGPA; Model 3b: γ_{10}=UGPA, γ_{20}=GRE V + Q + AW; Model 3c: γ_{10}=UGPA, γ_{20}=GRE V + Q; Model 3d: γ_{10}=UGPA, γ_{20}=GRE-Q, γ_{30}=GRE-V, γ_{40}=GRE-AW; Model 3e: γ_{10}=UGPA, γ_{20}=LSAT. The reduction in within level variance was computed based on the nonrounded values of τ_{00} and σ^2. Hence, although the values reported of τ_{00} and σ^2 across models are identical in the table, the reduction in variance was different.

$^{***}p < .001$, $^{**}p < .01$, $^{*}p < .05$

Table C4 Cross-Level Interactions Between Selectivity of Institutions (Median LSAT) and Level 1 Predictors

	Model			
	4a	4b	4c	4d
Fixed effects				
γ_{00}	3.23 (0.02)***	3.23 (0.02)***	3.23 (0.02)***	3.23 (0.02)***
γ_{10}	0.20 (0.04)***	0.22 (0.04)***	0.21 (0.04)***	0.23 (0.04)***
γ_{20}	0.07 (0.00)***	0.09 (0.01)***	0.01 (0.00)***	0.02 (0.00)***
γ_{30}			0.01 (0.00)***	
γ_{40}			0.05 (0.01)***	
γ_{01}	0.03 (0.00)***	0.03 (0.00)***	0.03 (0.00)***	0.03 (0.00)***
γ_{21}	−0.003 (0.00)***	−0.004 (0.00)***	−0.001 (0.00)**	−0.001 (0.00)***
γ_{31}			0.000 (0.00)	
γ_{41}			−0.001 (0.00)	
Variance components				
τ_{00}	0.01 (0.00)***	0.01 (0.00)***	0.01 (0.00)***	0.01 (0.00)***
τ_{11}	0.02 (0.02)	0.02 (0.01)	0.02 (0.02)	0.02 (0.01)
τ_{22}	—	0.00 (0.00)	0.00 (0.00)	0.00 (0.00)
τ_{33}			—	
τ_{44}			—	
σ^2	0.11 (0.00)***	0.11 (0.00)***	0.11 (0.00)***	0.11 (0.00)***
	vs. *Model 3b*	vs. *Model 3c*	vs. *Model 3d*	vs. *Model 3e*
Reduction in within level variance	0.36%	0%	0%	1.27%
Reduction in between level variance	80.06%	80.08%	80.04%	80.04%
Model fit				
-2LogLikelihood	1,065.5	1,071.5	1,049.9	1,050.2
	vs. *Model 3b*	vs. *Model 3c*	vs. *Model 3d*	vs. *Model 3e*
Δ Deviance	47.2(2)***	46.4(2)***	46.7(4)***	46.4(2)***

Note. Standard errors are shown in parenthesis. Parameters per model: Model 4a: γ_{10}=UGPA, γ_{20}=GREQVAW, γ_{01}=Median LSAT, γ_{21}=GRE V + Q + AW*Median LSAT; Model 4b: γ_{10}=UGPA, γ_{20}=GRE V + Q, γ_{01}=Median LSAT, γ_{21}=GREV+Q*Median LSAT; Model 4c: γ_{10}=UGPA, γ_{20}=GRE-Q, γ_{30}=GRE-V, γ_{40}=GRE-AW, γ_{01}=Median LSAT, γ_{21}=GRE-Q*Median LSAT, γ_{31}=GRE-V*Median LSAT, γ_{41}=GRE-AW*Median LSAT; Model 4d: γ_{10}=UGPA, γ_{20}=LSAT, γ_{01}=Median LSAT, γ_{21}=LSAT*Median LSAT. The reduction in within-level variance was computed based on the nonrounded values of τ_{00} and σ^2. Hence, although the values reported of τ_{00} and σ^2 across models are identical in the table, the reduction in variance was different.
$^{***}p < .001$, $^{**}p < .01$, $^{*}p < .05$

Results from Model 3a show UGPA as a Level 1 predictor explained 6.78% of within-school variance (Table C3). Models 3b through 3e revealed that GRE V + Q + AW, GRE V + Q, all three GRE sections together, and LSAT explained between 10% and 15% of the within-institution (Level 1) variance over and above the portion explained by UGPA, showing the added predictive validity of the GRE scores and LSAT. It should be noted that Models 3b through 3e explained very similar percentages of LGPA variance within schools, demonstrating the similar predictive strength of the sets of independent variables examined in these models.

RQ4. Is the relationship between GRE and LGPA moderated by institutional selectivity? (And for additional context, is the relationship between LSAT and LGPA moderated by institutional selectivity?)

Models 4c, 5c, and 6c resulted in a nonpositive definite matrix due to the estimates of τ_{33} and τ_{44}, indicating that there was no variability in the slopes of GRE-V and GRE-AW across institutions; hence, τ_{33} and τ_{44} were fixed to zero.

The results indicate that the models that included the interaction between Level 1 variables and institutional selectivity measured as the median institutional LSAT (Models 4a–4d), showed a significantly better fit to the data than the models without institutional selectivity (Models 3b–3e) according to the difference in deviance (Table C4). Although we found a significant interaction between institutional selectivity and Level 1 predictors, compared to models that did not include selectivity, Models 3b, 3c, and 3d did not explain additional within-school (Level 1) variance, while Model 4a explained only an additional 1.27%. Although these models did not account for within-school variance, they accounted for 80% of the between-institution variance in LGPA. In other words, selectivity measured as median institutional LSAT accounted for a large majority of variability in median LGPA across institutions, but the relationship between GRE scores and LGPA, and between student LSAT scores and LGPA, was very consistent across institutions regardless of selectivity.

Table C5 Cross-Level Interactions Between Selectivity (Dummy Coded) and Level 1 Variables

	Model							
	5a	5b	5c	5d	6a	6b	6c	6d
Fixed effects								
γ_{00}	3.41 (0.05)***	3.41 (0.05)***	3.41 (0.05)***	3.41 (0.05)***	2.96 (0.05)***	2.96 (0.05)***	2.96 (0.05)***	2.96 (0.05)***
γ_{10}	0.20 (0.04)***	0.22 (0.04)***	0.21 (0.05)***	0.23 (0.04)***	0.20 (0.04)***	0.22 (0.04)***	0.21 (0.05)***	0.23 (0.04)***
γ_{20}	0.05 (0.01)***	0.07 (0.02)***	0.01 (0.00)***	0.02 (0.00)***	0.09 (0.01)***	0.13 (0.01)***	0.02 (0.00)***	0.04 (0.00)***
γ_{30}			0.01 (0.00)**				0.01 (0.00)***	
γ_{40}			0.03 (0.02)				0.06 (0.02)***	
γ_{01}	−0.10 (0.07)	−0.10 (0.07)	−0.10 (0.07)	−0.10 (0.07)	0.35 (0.07)***	0.35 (0.07)***	0.35 (0.07)***	0.35 (0.07)***
γ_{02}	−0.45 (0.07)***	−0.45 (0.07)***	−0.45 (0.07)***	−0.45 (0.07)***	0.45 (0.07)***	0.45 (0.07)***	0.45 (0.07)***	0.45 (0.07)***
γ_{21}	0.01 (0.01)	0.00 (0.02)	0.00 (0.00)	0.00 (0.00)	−0.03 (0.01)*	−0.06 (0.02)**	−0.01 (0.00)*	−0.02 (0.00)**
γ_{31}			0.00 (0.00)				−0.01 (0.00)	
γ_{41}			0.04 (0.03)				0.01 (0.03)	
γ_{22}	0.04 (0.01)**	0.06 (0.02)*	0.01 (0.00)	0.02 (0.00)***	−0.04 (0.01)**	−0.06 (0.02)*	−0.01 (0.00)	−0.02 (0.00)***
γ_{32}			0.00 (0.00)				0.00 (0.00)	
γ_{42}			0.03 (0.03)				−0.03 (0.03)	
Variance components								
τ_{00}	0.02 (0.00)***	0.02 (0.01)***	0.02 (0.00)***	0.02 (0.01)***	0.02 (0.01)***	0.02 (0.01)**	0.02 (0.00)***	0.02 (0.01)***
τ_{11}	0.02 (0.02)	0.02 (0.02)	0.03 (0.02)	0.02 (0.01)*	0.02 (0.02)	0.02 (0.01)	0.03 (0.02)	0.02 (0.01)*
τ_{22}	0.00 (0.00)	0.00 (0.00)	0.00 (0.00)	0.00 (0.00)	0.00 (0.00)	0.00 (0.00)	0.00 (0.00)	0.00 (0.00)
τ_{33}			—				—	
τ_{44}			—				—	
σ^2	0.11 (0.00)***	0.11 (0.00)***	0.11 (0.00)***	0.11 (0.00)***	0.11 (0.00)***	0.11 (0.00)***	0.11 (0.00)***	0.11 (0.00)***
	vs. *3b*	vs. *3c*	vs. *3d*	vs. *3e*	vs. *3b*	vs. *3c*	vs. *3d*	vs. *3e*
Reduction L-1 variance	0%	0.09%	0.28%	1.45%	0%	0.09%	0.28%	1.45%
Reduction L-2 variance	70.22%	70.24%	70.19%	70.20%	70.22%	70.24%	70.19%	70.20%
Model fit								
-2Log Likelihood	1,080.1	1,085.1	1,061.6	1,060.7	1,080.1	1,085.1	1,061.6	1,060.7
	vs. *3b*	vs. *3c*	vs. *3d*	vs. *3e*	vs. *3b*	vs. *3c*	vs. *3d*	vs. *3e*
Δ Deviance	32.6(4)***	32.8(4)***	35(8)***	35.9(4)***	32.6(4)***	32.8(4)***	35(8)***	35.9(4)***

Note. L = level. Standard errors are shown in parenthesis. Predictors per model: Model 5a and 6a: γ_{10}=UGPA, γ_{20}=GRE V + Q + AW, γ_{01}=Dummy1, γ_{02}=Dummy2, γ_{21}=GRE V + Q + AW*Dummy1, γ_{22}=GRE V + Q + AW*Dummy2; Model 5b and 6b: γ_{10}=UGPA, γ_{20}=GRE V + Q, γ_{01}=Dummy1, γ_{02}=Dummy2, γ_{21}=GRE V + Q*Dummy1, γ_{22}=GRE V + Q*Dummy2; Model 5c and 6c: γ_{10}=UGPA, γ_{20}=GRE-Q, γ_{30}=GRE-V, γ_{40}=GRE-AW, γ_{01}=Dummy1, γ_{02}=Dummy2, γ_{21}=GRE-Q*Dummy1, γ_{31}=GRE-V*Dummy1, γ_{41}=GRE-AW*Dummy1, γ_{22}=GRE-Q*Dummy2, γ_{32}=GRE-V*Dummy2, γ_{42}=GRE-AW*Dummy2; Models 5d and 6d: γ_{10}=UGPA, γ_{20}=LSAT, γ_{01}=Dummy1, γ_{02}=Dummy2, γ_{21}=LSAT*Dummy1, γ_{22}=LSAT*Dummy2.
The reduction in within-level variance was computed based on the nonrounded values of τ_{00} and σ^2. $^{***}p < .001$, $^{**}p < .01$, $^{*}p < .05$

A similar pattern of results was found in the models that included two-way cross-level interactions between Level 1 predictors and selectivity broken into Groups I through III (Table C5). Although these models showed a significantly better fit to the data than the models that did not include selectivity as a Level 2 variable (Models 3b–3c), the addition of selectivity broken into three groups only explains an additional 1% to 3% of the within school variance. These models explain 70% of between school variance.

Results for Models 5a through 5d (where high selectivity served as the reference group) showed no significant interactions between high versus moderate selectivity and any Level 1 predictor (GRE V + Q + AW, GRE V + Q, each GRE section, and LSAT). We found a significant interaction between high versus low selectivity and GRE V + Q + AW (Model 5a), GRE V + Q (Model 5b), and LSAT (Model 5d). These results indicated that a stronger relationship between LGPA and each of the above-mentioned predictors was observed among students attending less selective institutions compared with their peers attending highly selective institutions.

When the institutions exhibiting low selectivity were modeled as the reference group (Models 6a–6d in Table C5), two-way cross-level interactions between the GRE measures and selectivity and the LSAT and selectivity showed a similar pattern as in Models 5a through 5d, with the addition of significant interactions between low versus medium selectivity and GRE V + Q + AW, GRE V + Q, and LSAT. These results indicated that a stronger relationship between LGPA and each of the above-mentioned predictors was found among students attending less selective institutions compared with their peers attending institutions exhibiting moderate selectivity. Although we observed statistically significant interactions, the small additional within-school variance explained (between 1% and 3%) indicates that the interactions are not practically meaningful.

Discussion and Conclusions for the Hierarchical Linear Modeling Analyses

We conducted a series of analyses to explore the validity of the GRE section scores and composites (and LSAT) for predicting LGPA, after controlling for the institutional variability. Overall, we found a positive relationship between GRE and LGPA (and between LSAT and LGPA) such that students with higher GRE scores (or with higher LSAT scores) tend to obtain higher LGPAs. The results also indicate that there does not seem to be any practical difference in the independent predictive validity of GRE and LSAT scores. Even after taking into account UGPA, GRE and LSAT explain approximately the same amount of additional variance in LGPA.

Additionally, we were interested in examining if the predictive validity of the GRE (and LSAT) varied depending on the selectivity of the law schools that participated. We found that selectivity measured as median institutional LSAT was a statistically significant moderator in the relationship between GRE and LGPA (and in the relationship between LSAT and LGPA). The results of the models that used selectivity as a categorical variable indicate that the relationship between GRE and LGPA (and between LSAT and LGPA) was stronger in less selective institutions than in moderately and highly selective institutions. We did not observe differences between highly versus moderately selective institutions. Although the models with selectivity as a moderator explained between 70% and 80% of the variance between law schools, they explained only between 1% and 3% of the within school variability. These results indicate that although there is variability across schools in terms of their selectivity, this variability does not have a practically significant impact on the validity of GRE (and LSAT).

Suggested citation:

Klieger, D. M., Bridgeman, B., Tannenbaum, R. J., Cline, F. A., & Olivera-Aguilar, M. (2018). *The validity of* GRE® *General Test scores for predicting academic performance at U.S. law schools* (Research Report No. RR-18-26). Princeton, NJ: Educational Testing Service. https://doi.org/10.1002/ets2.12213

Action Editor: Rebecca Zwick

Reviewers: Michael Kane and Daniel McCaffrey