INTRODUCTION

Recent studies have questioned the value of Graduate Record Exam (GRE) scores in predicting success in biomedical doctoral programs (Weiner 2014, Hall 2017, Moneta-Kohler 2017, Sealy 2019). Weiner (2014) and Hall (2017) also examined the predictive value of undergraduate GPA and found no correlation with subsequent research productivity. Moreover, reliance on GRE scores likely reinforce barriers against women, historically marginalized racial and ethnic populations, and those with fewer economic resources (FIND). Many institutions, including our own, have scrutinized their graduate admission process in the past few years, especially the GRE as an option or requirement (FIND).

Our program, the Master of Environmental Studies (MES) program at The Evergreen State College, required the GRE general exam for decades. Our applicants for this highly interdisciplinary program come from diverse disciplines and undergraduate institutions. GRE scores provided one common metric of academic aptitude and preparation in an application that also included undergraduate transcripts, statement of purpose, resume, three letters of recommendation, and other optional supporting materials. In our holistic review process, multiple MES faculty members review each student’s application for admission decisions and merit scholarship awards. Our primary consideration is whether a student will be able to succeed in a graduate program that includes core principles of environmental disciplines, integration across disciplines, quantitative research and statistical analysis, argumentative and technical writing, and an original research thesis. We seek to avoid admitting students who will spend money, effort, and time on a program for which they lack adequate preparation.

[Our aim fits with Briel et al (1993) statement that GRE was specifically designed to measure “basic developed abilities relevant to performance in graduate studies.” (Quoted in Kuncel 2001, p. 162.)

Given our concerns about equity and economic burden, we decided to evaluate the GRE’s predictive validity—whether the test provides reliable insight into an applicant’s future academic success in our program. We were interested in questions along the following lines:

* How well do GRE-Verbal scores predict meaningful participation in graduate seminars?
* How well do GRE-Quantitative scores predict understanding of research methodology and advanced statistics?
* How well do GRE-Writing scores predict success in a) argumentative essays, b) research papers, and c) original research theses?

We found that this scholarship, for the most part, provides limited insight about the questions that were most important to our admissions process. At the same time, we discovered that Evergreen’s practice of narrative evaluations, which document each student’s work for every course, provides granular and nuanced information about academic performance missing from most GRE predictive validity studies. We describe specific areas of contribution following an overview of this scholarship.

According to Kuncel et al (2001), validity studies since the the GRE’s development in the 1940s were “inconsistent and controversial” (p. 163). Validity studies completed during the past 20 years continue to offer ambiguous and conflicting evidence for graduate programs rethinking their admissions process.

On the one hand, multiple studies conclude that GRE scores provide moderate to strong predictive validity of student success in a variety of graduate and professional programs. In a comprehensive meta-analysis across multiple disciplines, Kuncel et al (2001) found that GRE scores and UGPA were “generalizably valid predictors of graduate grade point average, 1st-year graduate GPA, comprehensive examination scores, publication citation counts, and faculty ratings.” Specifically, GRE-Verbal and GRE-Quantitative for GGPA had greater predictive value (correlation RHO of .34 and .36, respectively) than undergraduate GPA (correlation RHO of .30), with even greater differences for graduate programs in the humanities, social sciences, and life sciences. (Analytical Writing component was introduced in XX.) In a subsequent analysis, Kuncel et al (2010) found that GRE-Verbal and GRE-Quantitative had predictive validity (RHO between .27 and .38) for both masters and doctoral programs across many disciplines. [Should I report correlation or RHO?] Kuncel et al (2010) called for additional primary research for “specific populations and situations” to support more robust meta-analyses, as well as the importance of using “other aspects of student performance.” Kuncel et al. found, interestingly, substantially greater variability in GGPA for masters than doctoral programs (maybe put below in discussion of restricted ranges).

Zero-order correlations and alternative metrics to zero-order correlations demonstrate that the GRE sections, includ-

ing the Analytical Writing section introduced in October 2002, provide value alone and in comparison to UGPA, the

predictor with which GRE sections are oen used by admissions committees in a compensatory manner.

9

Whether one

looks at the multiple correlation for regression models of GRE sections (R), the incremental coecient of determina-

tion (ΔR

2

) for GRE sections over and above UGPA, usefulness weight percentages (%C

xj

), or GRE quartile comparisons

(high/low or low/high), one observes abundant empirical evidence that the GRE sections provide value in decision mak-

ing if the accurate prediction of GGPA is one’s objective. GGPA remains the most common measure of graduate student

performance.

In addition, w

The Educational Testing Service, which owns and administers the GRE, has conducted multiple studies since the addition of the Analytical Writing component that support GRE predictive validity for all three components. Klieger et al (2014) analyzed five years of data on masters and doctoral programs at ten Florida universities using zero-order correlation and several multivariate analyses. For every analysis, they conclude, “one observes abundant empirical evidence that the GRE sections provide value in decision-making if the accurate prediction of GGPA is one’s objection.” (p36) They also predict that “with an outcome measure of academic achievement more reliable than GGPA, we would have seen higher values for validity.” (p36) ETS-sponsored research also found correlation between GRE scores and first-year and cumulative scores in 12 MBA programs (XX) and between GRE scores and first-year GPA in 21 law schools (XX). In its case for the GRE’s continuing relevance, “GRE Test Validity: Putting it in Perspective,” ETS argues that correlation values between 0.22 and 0.37 are regularly used to inform decisions in public health, medical treatment, and psychology (ETS website). We agree that such correlation values, if substantiated by our own assessment, would be useful for graduate admissions and merit awards, especially when GRE scores are used as one point of triangulation along with other values. [What, if anything does ETS say about correlation values for gender, race, etc? I can’t remember seeing anything!] Notably, ETS does not include correlation values for gender, race/ethnicity, or parental income in this chart. Such values have become a major concern for graduate programs seeking better diversity, equity, and inclusion.

On the other hand, several recent studies have undermined the predictive validity of GRE general test scores for doctoral programs, especially in the biomedical sciences. A study of 280 graduate students in biomedical science programs at the University of North Carolina—Chapel Hill found “no correlation of test scores, grades, amount of previous research experience, or faculty interview ratings with high or low productivity”; instead categorical ratings from recommendation letter writers was the sole application material that showed strong correlation with productivity (Hall 2017). A study of 683 graduate students in Vanderbilt University’s biomedical science programs concluded that GRE scores did not usefully predict “who will graduate with a Ph.D., pass the qualifying exam, have a shorter time to defence, deliver more conference presentations, publish more first author papers, or obtain an individual grant or fellowship” (Moneta-Kohler 2017). Other studies of doctoral programs in biomedicine have reached similar conclusions (Werner 2014, Sealy 2019). [LOOK FOR OTHER STUDIES]

Our impression, after reviewing GRE validity studies from the past decade, is that GRE scores have some predictive power for GGPA across many types and levels of graduate programs, which varies significantly across these programs. Specifically, GRE scores appear likely to have less predictive validity for success in doctoral programs with low GGPA variability and greater emphasis on research productivity. What does that mean for our program? MES might be considered a hybrid program with elements of academic and professional programs. Because MES requires a research thesis for graduation, the evidence from doctoral programs and research productivity is highly relevant. At the same time, the vast majority of our students do not pursue doctoral programs or academic research careers. Given the lack of clear guidance, we decided to conduct our own validity study. This article responds to the call for studies on the predictive validity of the GRE, as well as other application materials, in diverse graduate programs. [For example, there are XX masters programs in interdisciplinary environmental studies, sustainability, etc.] In addition, this study contributes to three underexamined areas of GRE validity studies.

One underexamined area is that most studies use datasets with restricted ranges in both application and outcome measures. Highly selective graduate programs typically admit students with high GRE scores and thus provide limited insight on how students across the spectrum of GRE scores perform in graduate programs. One exception is Sealy (2019) who assessed graduate performance of 32 students from historically underrepresented populations selected through a holistic admissions process. Evergreen’s MES program has enrolled students with a broad range of GRE scores. Our program is moderately selective. Our evaluation of applications focuses on whether students can successfully complete a rigorous academic program rather than on selecting the most elite cohort of graduate students. We have used a holistic assessment of applications throughout the admissions process, which means that we have no minimum GRE requirement and did not prescreen applications on GRE or GPA. In practice, we used the GRE as one point of information that, along with other points, to help us triangulate evidence showing student preparation. For instance, when considering whether a student has adequate analytical and quantitative foundations for Research Design and Quantitative Methods, we considered coursework and grades in quantitative subjects, work or volunteer experience, letters of recommendation, and GRE-Quantitative scores. A student might demonstrate adequate aptitude or preparation through one or, ideally, several of these application components. [Calculate range, mean, standard deviation for each GRE component and UGPA]

A second underexamined area of most GRE validity studies is the use of relatively blunt outcome measures such as first-year GPA or comprehensive GPA. As mentioned, existing scholarship provides limited guidance about specific questions we faced in our admissions process. Most GRE validity studies, especially for masters students, use blunt measures of graduate performance such as graduate GPAs (first-year or comprehensive) or degree completion as their primary outcomes. Such measures provide minimal insight into specific academic capabilities or growth areas that might hinder student academic success.

In addition, many graduate programs, especially at the doctoral level, have restricted range in GPAs as the result of grade inflation. Kuncel et al (2010) found that doctoral programs had significantly less variance in graduate GPA than masters programs [DATA]. Evergreen’s unique pedagogical and evaluation model, in this case, provides data and insight that might be of broader interest to many graduate programs considering whether and how to use GRE scores in their admissions process. All Evergreen courses, including those in the three graduate programs, use narrative evaluations to document each student’s competencies and accomplishments in specific academic skills. As a consequence of these detailed narrative evaluations, we can examine the GRE’s predictive validity using more nuanced outcome measures. Specifically, narrative evaluations for first-year MES core programs (which contain 24 of the required 72 credits for graduation) assess the quality of seminar participation, analytical essay writing, literature research and technical writing, public speaking, statistical analysis, collaboration and teamwork on group project, and project management. Some evaluations also provide insight on work ethic, persistence, adaptability, or other qualities. Our narrative evaluations thus provide more nuanced evidence about student achievement than a highly distilled “catch all” metric like graduate GPA.

A third underexamined area concerns the predictive longevity of application materials like UGPA and GRE scores. We hypothesize that the GRE will have declining predictive value over a student’s time in the program, especially if the program is effectively educating students. Many studies contain elements of longitudinal trends (e.g. first-year GPA vs. comprehensive GPA) but lack more granular assessment of outcomes over the course of the first year when students are making the transition to graduate-level work. Our narrative evaluations for the first two quarters contain a paragraph for each student that assesses 1) seminar contributions, 2) analytical essays, and 3) research papers. Evaluations for all three quarters contain an assessment of overall performance. These components provide longitudinal data for each student that allows us to compare UGPA and GRE validity over the course of the first year.

METHODS

**Academic Program and Student Population**

Evergreen established a Master of Environmental Studies (MES) Program in 1986. With incoming cohorts of 40-50 each fall, MES is one of the oldest and largest of such programs. Since its inception, MES adopted key elements of the Evergreen educational model: a core curriculum comprised of interdisciplinary team-taught courses, electives and internships, and an original research thesis; a cohort model and intentional development of learning communities; and assessment of student work through narrative evaluations and end-of-quarter evaluation conferences. These evaluations comprise the transcript, which does not include grades or other numerical scores describing overall performance. (JESS article, MES website)

MES integrates elements of both academic and professional programs. On the one hand, our courses emphasize natural and social sciences, environmental policy, and research methodology. Students must complete a thesis based on original research in order to graduate. On the other hand, our program incorporates the development of practical skills (including optional certification in Geographical Information Systems) and extensive internship opportunities. While several students pursue doctoral degrees after earning their MES degree, most enter the workforce as environmental professionals.

MES student profile? Demographic breakdown?

This study is based on admission materials and evaluation scores for MES students who began our program in 2014, 2015, and 2016. In the Fall 2014 cohort, 43 students completed the first quarter, 41 students completed the second quarter, and 40 students completed the third quarter of the core sequence. In the Fall 2015 cohort, 43 students completed the first quarter, 39 students completed the second quarter, and 39 students completed the third quarter of the core sequence. In the Fall 2016 cohort, 45 students completed the first quarter, 43 students completed the second quarter, and 42 students completed the third quarter of the core sequence.

**Application Materials (GPA, GRE)**

We collected the followingdata as application requirements for these cohorts:GRE general exam scores, undergraduate transcripts, statement of purpose, resume, and three letters of recommendation. Based on these materials, five quantitative values inform our holistic application review: GRE-Verbal, GRE-Quantitative, GRE-Analytical Writing, GPA-Total (based on the entire undergraduate record), GPA-60/90 (based on the last 60 semester credits or last 90 quarter credits). We looked for correlation between each of these values and the transcript scores. Two caveats are important. First, our application policy does not require applicants with a previous masters degree to submit GRE scores, thus 6 of 131 students were not included in the GRE validity analysis. Second, many MES students completed their undergraduate degrees at Evergreen or other institutions with narrative evaluations, thus 35 of 131 (N=97, I’m one off below) students were not included in the UGPA validity analysis. [Higher than it should be? Edwards (alternative Florida college?), Horkings-Brigham, Keon, McCormick, Pecor, Thurston)

**Transcript Scoring and Inter Rater Reliability**

Three Evergreen students (two undergraduate, one graduate) scored narrative evaluations for three 8-credit courses of the first-year core curriculum. Following an approved Human Subjects Research process, all identifiable personal information was redacted prior to this scoring exercise. Academic performance on key activities or individual assignments, as well as overall performance, was scored on a 10-point scale (Table X; Figure X—example rubric?). Each evaluation was also given an effort score on a 7-point scale (Figure Y—example rubric?). We refined the scoring rubric following a preliminary scoring exercise on a previous cohort’s evaluations.

|  |  |  |
| --- | --- | --- |
| **Conceptualizing our Regional Environment (Fall)** | **Ecological and Social Sustainability (Winter)** | **Research Design and Quantitative Methods (Spring)** |
| Seminar Contributions | Seminar Contributions | Exams and Labs |
| Analytical Essays | Analytical Essays | Overall Performance |
| Research Paper | Research Paper | Effort/Work Ethic/Perseverance |
| Overall Performance | Overall Performance |  |
| Effort/Work Ethic/Perseverance | Effort/Work Ethic/Perseverance |  |

Intra rater reliability was assessed using intraclass correlation (TK Koo and MY Li 2016). Intraclass correlation (ICC) values and their 95% confidence intervals were calculated using SPSS statistical package version XX (SPSS Inc, Chicago, IL) based on a mean-rating (k=3), absolute-agreement, 2-way mixed-effects model. We identified reliability level based on the following scale: poor reliability (ICC < 0.5), moderate reliability (0.5 < ICC < 0.75), good reliability (0.75 < ICC < 0.9), excellent reliability (0.9 < ICC). The general pattern was that specific assignments (essays, research papers, exams and labs) had excellent reliability, seminar contributions and overall performance had good reliability, and effort/perseverance had moderate reliability.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Evaluation Component (Quarter)** | **Intraclass Correlation (ICC Value)** | **95% Confidence Interval** | **Level of Reliability (ICC Value)** | **Level of Reliability (95% Confidence Interval)** |
| Seminar (Fall) | 0.855 | 0.753-0.909 | Good | Good |
| Analytical Essays (Fall) | 0.926 | 0.900-0.945 | Excellent | Excellent |
| Research Paper (Fall) | 0.919 | 0.890-0.940 | Excellent | Good |
| Overall (Fall) | 0.882 | 0.833-0.917 | Good | Good |
| Effort/Perseverance (Fall) | 0.513 | 0.325-0.652 | Moderate | Poor |
| Seminar (Winter) | 0.826 | 0.813-0.901 | Good | Good |
| Analytical Essays (Winter) | 0.917 | 0.888-0.940 | Excellent | Good |
| Research Paper (Winter) | 0.916 | 0.886-0.939 | Excellent | Good |
| Overall (Winter) | 0.903 | 0.868-0.929 | Excellent | Good |
| Effort/Perseverance (Winter) | 0.677 | 0.537-0.774 | Moderate | Moderate |
| Exams and Labs (Spring) | 0.909 | 0.846-0.943 | Excellent | Good |
| Overall (Spring) | 0.836 | 0.726-0.897 | Good | Moderate |
| Effort/Perseverance (Spring) | 0.656 | 0.529-0.752 | Moderate | Moderate |

**RESULTS AND DISCUSSION**

**Summary Statistics of UGPA and GRE Scores**

Most UGPA and GRE validity studies have UGPA and GRE datasets with restricted ranges because of the widespread practice of screening out applications below certain UGPAs and GRE scores. Such practices make it difficult to evaluate the predictive validity of UGPA and GRE scores (Kuncel et al 2010???). Because our program has never used this screening method, so the undergraduate GPA and GRE scores (verbal, quantitative, and writing) of our study participants allow for meaningful analysis across the spectrum of potential graduate students. Study participants had a mean UGPA of 3.395 (SD=0.321, N=97) and broad range (minimum=2.39, maximum=4.0). Study participants had a mean verbal percentile of 66.464 (SD=21.803, N=125) and broad range (minimum=13, maximum= 99). Study participants had a mean quantitative percentile of 39.192 (SD=21.613, N=125) and broad range (minimum=2, maximum=97). Study participants had a mean writing percentile of 51.824 (SD=23.614) and broad range (minimum=7, maximum=98). Our holistic admissions process thus provides a dataset for testing correlation of UGPA and GRE scores with graduate academic performance across a more diverse range of students than most previous studies (see Sealy 2019 as one exception).

**Summary Statistics for Evaluation Scoring**

Most UGPA and GRE validity studies also have restricted ranges when it comes to outcome measures because of the significant amount of grade inflation in many graduate programs, especially those leading to Ph.D. degrees (Kuncel et al 2010???) Our scoring rubric, with a 10-point scale for academic performance and a 7-point scale for effort, was intended to provide greater variation in outcome measures associated with specific skills and overall performance in each course. We succeeded in creating enough variation to provide statistically significant results in correlation analysis for multiple outcome variables. Table XX provides summary statistics for academic performance categories. Note that these are mean scores for three raters, so a maximum of 9.67 represents one rater score of 9 and two rater scores of 10. Note that Effort scores were on a -3 to +3 scale.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Evaluation Component** | **Number of Students** | **Mean** | **Standard Deviation** | **Minimum** | **Maximum** |
| Quarter 1 Seminar | 131 | 7.04 | 1.16 | 4.67 | 9.67 |
| Quarter 1 Essays | 131 | 6.97 | 1.33 | 3 | 10 |
| Quarter 1 Research Paper | 130 | 7.58 | 1.45 | 2.67 | 10 |
| Quarter 1 Overall Performance | 131 | 7.42 | 0.98 | 3.67 | 9.67 |
| Quarter 1 Effort | 131 | 0.81 | 0.63 | -1.33 | 2.33 |
| Quarter 2 Seminar | 124 | 7.55 | 1.43 | 3 | 9.67 |
| Quarter 2 Essays | 124 | 6.62 | 1.54 | 2.33 | 10 |
| Quarter 2 Research Paper | 124 | 7.28 | 1.59 | 2 | 10 |
| Quarter 2 Overall Performance | 124 | 7.20 | 1.24 | 4 | 9.33 |
| Quarter 2 Effort | 124 | 0.94 | 0.845 | -1.67 | 2.67 |
| Quarter 3 Exams/Labs | 122 | 6.33 | 1.70 | 2.33 | 9.67 |
| Quarter 3 Overall Performance | 122 | 6.64 | 1.38 | 2.33 | 9.67 |
| Quarter 3 Effort | 122 | 0.66 | 0.72 | -1.33 | 3 |

**Correlation Matrix**

Table XX provides the correlation values and probabilities for our study. Color code for 95% confidence (p=0.05) and 90% confidence (p=0.10)? We present the entire correlation matrix, including values that are not statistically significant, and consider all correlation values as providing meaningful information for graduate admissions. Whereas some studies simply report the lack of correlation between UGPA or GRE scores and outcome variables, we believe the entire set of these results are relevant for our questions about UGPA and GRE validity. Our results showed that 14 of 52 correlation values were statistically significant (p=0.05) prior to the Bonferroni correction, including 2 of 14 for UGPA and 12 of 14 for GRE scores. They also show a clear pattern of higher correlation values linked to statistical significance. All correlation values 0.2226 or greater were statistically significant and all correlation values 0.2124 or lesser were not statistically significant. In other words, the statistical significance of positive correlation is a reflection of correlation strength rather than sample size or other factors. For an assessment of relative correlation between UGPA and GRE, we think it is important to consider the entire correlation matrix.

**We also conducted a Bonferonni correction…x of 14 for UGPA were significant, x of 39 for GRE were significant. [Should I be including effort in this analysis? Maybe discuss in a separate section?]**

Multivariate Correlations

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | **SEM 1** | **ESSAY 1 MEAN** | **TECH 1 MEAN** | **OVERALL 1 MEAN** | **EFFORT 1 MEAN** | **SEM 2 MEAN** | **ESSAY 2 MEAN** | **CAND 2 MEAN** | **OVERALL 2 MEAN** | **EFFORT 2 MEAN** | **EXAM 3 MEAN** | **OVERALL 3 MEAN** | **EFFORT 3 MEAN** |
| CUM GPA | 0.0316 | 0.1896 | 0.1650 | 0.2124 | -0.0327 | -0.0516 | 0.2226 | 0.1066 | 0.2328 | 0.0931 | 0.1639 | 0.1157 | -0.0869 |
| VERBAL % | 0.1743 | 0.2022 | 0.1745 | 0.2775 | 0.0848 | 0.2619 | 0.3038 | 0.3221 | 0.3672 | 0.0480 | 0.2870 | 0.2751 | 0.0830 |
| QUANT % | 0.1208 | 0.0230 | -0.0059 | 0.0785 | 0.0214 | 0.2912 | 0.1192 | 0.2627 | 0.2172 | 0.0401 | 0.1543 | 0.1333 | -0.1197 |
| WRITING % | 0.0560 | 0.0852 | 0.0319 | 0.1120 | -0.0147 | 0.0524 | 0.1330 | 0.3020 | 0.3200 | 0.1774 | 0.1489 | 0.0943 | 0.1296 |

Correlation Probability

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | **SEM 1 MEAN** | **ESSAY 1 MEAN** | **TECH 1 MEAN** | **OVERALL 1 MEAN** | **EFFORT 1 MEAN** | **SEM 2 MEAN** | **ESSAY 2 MEAN** | **CAND 2 MEAN** | **OVERALL 2 MEAN** | **EFFORT 2 MEAN** | **EXAM 3 MEAN** | **OVERALL 3 MEAN** | **EFFORT 3 MEAN** |
| CUM GPA | 0.7766 | 0.0860 | 0.1362 | 0.0539 | 0.7694 | 0.6429 | 0.0431 | 0.3374 | 0.0342 | 0.4026 | 0.1388 | 0.2974 | 0.4349 |
| VERBAL % | 0.1150 | 0.0668 | 0.1146 | 0.0111 | 0.4460 | 0.0168 | 0.0052 | 0.0030 | 0.0006 | 0.6668 | 0.0085 | 0.0118 | 0.4556 |
| QUANT % | 0.2768 | 0.8366 | 0.9580 | 0.4804 | 0.8479 | 0.0076 | 0.2831 | 0.0164 | 0.0486 | 0.7191 | 0.1636 | 0.2296 | 0.2812 |
| WRITING % | 0.6150 | 0.4440 | 0.7744 | 0.3132 | 0.8947 | 0.6380 | 0.2305 | 0.0055 | 0.0032 | 0.1086 | 0.1792 | 0.3965 | 0.2430 |

**Substantive results/discussion (finally!)**

UGPA: We found no statistically significant relationship (p=0.05 I’m guessing…after Bonferroni correction) between UGPA and academic performance. However, we found meaningful patterns (at p=0.10 I’m guessing etc) for UGPA and overall performance in the first two courses (r=.2124 and r=2328, respectively) of the core sequence.

Our results suggest that UGPA might provide some insight into graduate academic performance, but that this correlation is relatively weak and variable across different courses.

One implication is that UGPA should not be relied upon as a cutoff value or filter prior to a holistic evaluation of applications.

Why the relatively weak correlation? Variation in grading practices of different institutions and disciplines. (Source?)

Should we continue using UGPA and transcripts in the admissions process? Yes, but we should consider UGPA as a value within a greater context that considers our existing knowledge of the institution (e.g. academic rigor, grade inflation) and specific coursework. For instance, a more granular assessment of undergraduate transcripts often reveals patterns in grades that might predict performance in graduate environmental studies. What courses have students taken in particular areas (e.g. natural science, social science, humanities, art, or professional-oriented courses)? How did they perform in environmental coursework versus other coursework? How did their academic performance evolve over their college career? Our professional experience and judgment suggests the value of such considerations. This study does not provide supporting evidence, but we have begun collecting information from applicants about the pattern of their undergraduate coursework that will allow us to evaluate such factors in a future study.

GRE: Our results suggest that GRE scores provide moderately strong correlation for some coursework, as well as substantial variation in the correlation values across the spectrum of academic skills. The strongest correlations were for verbal scores in seminar performance, essays, research papers, and overall performance. Perhaps surprisingly, verbal scores predicated success in our research design and advanced statistics course much better than quantitative scores.

What are the implications?

NEW MATERIAL STOPS HERE!

**CONCLUSION**

Brief note on timeline…[where does it go?]

In 2017 we (KF and AA) completed a preliminary study to inform whether, given the various problems with bias, access, and validity, we should continue requiring the GRE. KF scored four years of evalutions for two first-year programs and KF and AA analyzed the results. Based on this initial assessment, we replaced the GRE requirement with an analytical essay focused on environmental policy beginning with the Fall 2018 cohort. For the reasons discussed, above, we decided that a more rigorous undertaking of this analysis (with multiple raters who had not taught each of these students and written some of their evaluations) might be valuable for other graduate programs evaluating the use of the GRE and considering potential alternatives. In the final section, we discuss our decision to eliminate the GRE and modify both our application requirements and holistic evaluation process in order to assess the impact of these modifications and modify, as necessary, our admissions process.

**NOTES**

Would we undermine our DEI goals by putting more emphasis on undergraduate college reputations and letters of recommendation, which might privilege—in their own way—students with historical privilege, greater economic resources, or more experience navigating the academia and professional social systems. Each of these application materials contain their own forms of bias (XX). Having more information, especially information that allowed us to compare students across many disciplines and institutions, seemed like it might allow us to make better decisions about student preparedness. With such concerns and questions, we turned to the literature on GRE and graduate admissions.

Something about how we are primarily interested in whether students have adequate preparation to succeed in graduate school.

Our own institution went through such introspection in 2017. The literature did not provide a clear direction…

Studies suggest a great deal of variability in terms of the value of GPA and GRE in different kinds and levels of programs.

The graduate admission process at many institutions is . In the past few years many graduate schools, programs, and departments the GRE the role of the Graduate Record Exam (GRE) in their application process. Many have eliminated it

Critique of GRE Validity

The GRE validity research offers conflicting evidence about whether the GRE accurately predicts success in graduate school and careers.

Studies over the past decade have raised concerns about 1) validity of GRE as a predictor of success in graduate school for all students, 2) bias in GRE testing that privileges Asian and white males over other testtakers, 3) use of GRE as filter or screen for graduate admissions that screens out a higher percentage of female and people of color.

“ETS” Studies showing GRE Validity

At the same time, the Educational Testing Service continues to emphasize the value of the GRE for graduate programs faced with difficult decisions about admission and support.

Shortcomings in Validity Research

Most research focuses on performance of students in top 50% of GRE scores since those are the ones who make it into grad school. Applicants who are not in top 50% just don’t make it through the first filtering process. [TESC, which has adopted holistic evaluation throughout its history, admits and enrolls students across the spectrum of GRE scores.]

Most research uses blunt outcome measures (GGPA, completion, etc.) that does not provide insight on whether the GRE measures specific abilities and skills that serve as an important foundation for success in graduate school.

What this study offers…

This study addresses some of the existing weaknesses in GRE validity studies by drawing upon the unique data provided by narrative evaluations.

Evergreen / MES / Narrative

d

Studies over the past decade have raised concerns about 1) validity of GRE as a predictor of success in graduate school for all students, 2) bias in GRE testing that privileges Asian and white males over other testtakers, 3) use of GRE as filter or screen for graduate admissions that screens out a higher percentage of female and people of color.

GRExit

GRExit is taking different forms. Some graduate programs are eliminating GRE scores from the application process entirely… Others are making scores optional…Others are requiring GRE scores but using them in an more explicitly holistic evaluation process (e.g., eliminating the GRE as part of the initial filtering or screening, following the ETS advice to use GRE as one piece of data.)

During this period, graduate programs must decide whether to accept GRE scores as an option or requirement. If they do accept/require, they must determine how to use GRE scores in the evaluation of applications. They face these decisions in the midst of a confusing, contested realm of evidence and advice about the validity of the GRE for predicting graduate school performance and the best way to incorporate GRE scores into admissions decisions.

Many grad programs are getting rid of their GRE requirement or making them optional. We are in the midst of a GRExit wave that builds upon itself as schools attempt to attract SOC and worry about losing students who might filter graduate programs based on whether they require the GRE.

Symbolic value has shifted from marker of excellence to marker of elitism and exclusion. I should find some good quotes from popular articles to illustrate this issue.

For programs there’s a very practical concern that some highly accomplished students, for economic and practical and philosophical reasons, may avoid programs that require GRE.

Promise of GRE and other standardized tests. Admissions based on aptitude and preparation rather than who you know.

Some grad programs who, for now, are keeping the GRE are adopting more holistic assessment of grad applications. This article contributes to validity literature (which is still contested) and how GRE might contribute to holistic assessment.

METHODS