The Future of Engineering Education

2024 Annual Conference & Exposition

Oregon Convention Center
Portland, OR . June 23 - 26, 2024

ASEE

Paper ID #44022

# The GRE in Admissions: Examining the Evidence and Arguments

**Dr. Edward F. Gehringer, North Carolina State University**

Dr. Gehringer is an associate professor in the Departments of Computer Science, and Electrical & Computer Engineering. His research interests include computerized assessment systems, and the use of natural-language processing to improve the quality of rev

# The GRE in Admissions: Examining the Evidence and Arguments

**Abstract**

*A recent trend in graduate admissions has been to eliminate the requirement to submit GRE scores (called "test optional"), or even to prohibit their use in admissions decisions. This paper summarizes the arguments for and against the use of standardized tests in general, and the GRE in particular. The GRE provides a comparison that is at least facially objective, though scores may be influenced by factors such as test anxiety. GRE scores seem to predict outcomes like GPA and degree completion, but different surveys and statistical methods lead to different conclusions. The GRE may enable programs to better target their admission offers to students who can succeed, but it may also discourage minority applications.*

## 1. Introduction

Every graduate program desires—or should desire—to admit only students who will be able to succeed in their course of study. At the same time, almost all programs are committed to broadening educational opportunity by admitting non-traditional students and seeing them through to graduation. Standardized tests such as the Graduate Record Exam were conceived as a way to discover talent that would not be apparent using traditional metrics such as prior educational background and grade-point averages. The GRE arose as a joint research project by Harvard, Yale, Columbia, and Princeton in 1936 [1]. The University of Wisconsin was the first outside school to use it, on an experimental basis, in 1938 [2]. In 1949, the GRE was acquired by the Educational Testing Service [1]. Its latest revision, in 2011, turned it into a fully adaptive test, meaning that a student's answers to initial questions were used to tune the difficulty of subsequent questions. It also allowed students to go back to questions they had skipped, or change their answers to previously completed questions [3].

The GRE's aim is honorable: to provide an objective estimate of a student's readiness for graduate study. But it has come under fire for a perceived tendency to decline or discourage applications from racial minorities and lower socio-economic status students. A number of research studies have examined questions like whether the GRE actually provides an objective comparison, whether students can get an unfair advantage from test-preparation services, or whether the GRE has predictive validity; that is, whether it can effectively forecast which students will complete the program they are seeking admission to. Because of a number of confounding factors, these questions are difficult to address directly, so studies differ markedly in their conclusions. This paper will endeavor to analyze some of the assumptions and results, and explain why observers draw such different conclusions.

The paper is organized as follows. Section 2 investigates whether the GRE provides an objective comparison of the applicants' abilities for graduate study. One issue that figures prominently in this discussion is cost—the claim that the GRE discriminates against poorer applicants. It merits its own analysis, in Section 2.1. A separate issue is predictive validity, which is taken up in Section 3. Given that schools desire to admit students who can succeed in their programs, an admission test can help if and only if it helps to separate those who can succeed from those who wouldn't. This question has been hotly debated, but ultimately, it comes down to whether the GRE adds predictive value to admissions decisions, not whether the GRE can be used to predict outcomes on its own. Section 4 describes the most prominent alternative to the use of standardized tests: holistic admissions. Regardless of how accurate metrics such as the GRE are, admissions decisions are about *individuals*, and no individual's competencies can be fully captured by a set of numbers. But is it practical to make effective comparisons of individuals apart from using a set of measures? Section 5 summarizes the issues covered in the paper, and provides a table of points and counterpoints.

## 2. Objectivity of Comparison

The goal of a standardized test is to provide an **objective** measure of a candidate's academic abilities. It attempts to provide a uniform way to compare candidates from various institutions, courses of study, and educational backgrounds. Few institutions, however, would use the GRE by itself to make admission decisions [4]. Standardized tests are useful, but they work best when combined with other indicators, including undergraduate grades. Exam results and academic transcripts together offer a more complete picture of a candidate's chances of succeeding in graduate school [5].

Standardized testing is intended to create a level playing field for all applicants. All candidates take the same exam in the same circumstances. This means that test results are less influenced by bias or personal ties, resulting in a more **meritocratic** selection process [6]. Standardized tests are administered according to established procedures, and results are assessed according to well defined criteria. The fact that both candidates and institutions are aware of how tests are scored and interpreted helps make the admissions process **transparent** and fair [7]. Test results can be used as a **benchmark** to compare an applicant's performance to national norms. This can assist in determining a candidate's standing in relation to others from very different backgrounds. For international applicants, standardized tests serve as a common metric that transcends differences in educational systems and grading scales.

On the other hand, the objectivity of standardized tests has been called into question for many reasons. One prominent reason is **cultural bias** [8]. This can potentially influence questions designed to measure intelligence; one example given by Kim and Zabelina [9] is a case where Westerners and Liberians were asked to sort 20 objects into categories. Westerners tended to group them into categories like foods, implements, containers, and clothing, whereas Kpelle people tended to group them by utility, placing a knife with a potato because the knife could be

used to cut the potato. **Socioeconomic bias** is another concern. Students who can afford test-prep services may have a significant advantage over those who cannot [10] (we return to this issue in Section 2.1). Some, such as the National Education Association, believe that the well documented and persistent tendency [11] of minority students to perform more poorly on these tests is evidence of **racism** [12]. They note that standardized testing rose to prominence in the eugenics era, when it was common to believe that intelligence was heavily influenced by race. The argument is not that these tests are *designed* to be biased, but just that the developers lack the cultural competence to understand different groups' knowledge base, and thus that the tests inadvertently exhibit bias against members of those groups.

**Language** is a hurdle for many test takers. Since the tests are administered in English, students from other linguistic backgrounds may not perform up to their ability, particularly on the verbal portion [13]. One recent study [14], though, challenges this assumption. Moreover, the tests are not designed to measure **soft skills** and leadership, which are important factors affecting success in graduate school [15]. It should be noted that the Educational Testing Service, which administers the GRE, is aware of this limitation and is working on another test to assess these attributes [16]. But it is not part of the current GRE.

**Test anxiety** [17] is another reason why standardized tests might not provide an objective comparison of candidates. Many students experience feelings of worry, nervousness, and self-doubt before or during an exam. If it affects their performance, then their test scores are not an authentic indication of their ability to succeed in a graduate program. One study of GRE test-takers in 1986 [18] showed that a higher degree of anxiety was associated with poorer performance on the test. A followup study using a computer-based version of the GRE failed to find that the handicap of test anxiety could be mitigated by electronic, rather than paper-based, administration [19]. Though anxiety in these studies was self-reported, it is corroborated by medical evidence. One study looked at test anxiety in 93 third- to eighth-grade students in a New Orleans charter school [20]. It found that students with the largest changes in cortisol on test day performed worse on standardized tests. The article claimed that test anxiety could reduce test scores by up to 12%, but it is not clear how this conclusion was arrived at.

Another concern is that **undergraduate institution** plays a role in determining GRE performance. Some research shows that students from selective undergraduate schools perform better than students from less selective schools [21]. This could be because those institutions offer more challenging courses, which better prepare students to excel on the GRE. However, this observation doesn't establish that the GRE is unfair. After all, students get into selective institutions by demonstrating aptitude, so they might be expected to exhibit greater aptitude for graduate study. To establish that a good undergraduate institution confers an unwarranted advantage, one would have to show that, once admitted to graduate school, students from selective undergraduate institutions perform below expectations relative to their peers from lower-ranked schools. And there does not seem to be any research addressing that question.

## 2.1 The issue of cost

Cost figures into the impact of the GRE in two separate ways. First, some students might be discouraged from applying because they can't afford the $205 fee—or because they can't afford to to take it multiple times in an effort to improve their score. The American Psychological Association reported that as schools dropped their GRE requirements, their applications for both masters and doctoral programs increased [22]. However, the data did not separate schools that continued to require the GRE from schools that dropped it, so it is not obvious that dropping the requirement was the reason that applications increased. At first glance, it might appear that the GRE is a minor expense compared with the expense of application fees to multiple institutions, and later, of graduate tuition. But that is not necessarily the case, because talented students may receive assistantship offers [23], which will pay their tuition when they matriculate. For such students, especially from poor countries or poor families, the cost of the test may be a major burden that can be averted if they simply avoid applying to programs that require the test.

The second reason that the GRE may discriminate socioecomically is because wealthier students can afford to take test-preparation programs to boost their score on the test. On this issue, much more research has been done on the college admissions tests than on the GRE, but the same considerations would seem to apply to both, so the SAT/ACT results may be applicable. Montgomery and Lilly [24] found an increase of 23.5 points on the SAT verbal and 32.7 points on the SAT math for students in test-preparation programs compared to those who weren't. Beginning preparation earlier and taking more practice tests helped students to improve their scores more [25]. The same effect has been observed for the ACT test [26]. Use of the ACT Online Prep program, where 6 practice sessions raised scores 1.22 points, and 16 practice sessions were associated with a 1.60-point increase [27].

As to research on the GRE itself, Swinton and Powers created a preparation course [28] for the analytical portion of the GRE, which was new at the time (1983). They found that, even controlling for other factors such as scores on the other GRE portions, the coached students scored 66 points higher on the 200–800 scale then in effect. As a result, the analytical portion of the test was redesigned, and the two "most susceptible" types of questions were removed [29].

Taken at face value, these findings would support a claim that the GRE is inequitable. But was it the prep course or the studying that raised scores? Powers and Swinton [30] repackaged their prep course for self-study. Students who used it in as little as four hours of self-study raised their scores by 53 points, or 80% as much as the coached students. The same phenomenon has been observed with the SAT. The College Board developed a booklet called "Taking the SAT" and, in a prepublication experiment, sent it to a random group of high-school juniors who had registered to take the test [29]. Those who simply *received the booklet* scored a significant 8 points higher on the math portion of the test (differences on the verbal portion were not significant).

It is by now well established that students can raise their scores by using effective study habits, which they can learn either from a test-preparation program or self-study materials that are widely available from the testing services' own web sites. This means that the advantage to wealthier students is limited, but obviously it still remains to some extent, because more people will persist at an activity if they have someone else supporting them in the endeavor than if they have to do it on their own.

Another implication is worth noting. The testing event itself helps students learn material. This is called the *testing effect* [31]. Candidates who study for an assessment will retain more knowledge than candidates who don't. To the extent that standardized-test requirements induce applicants to study, the tests themselves will help candidates succeed, relative to those who don't take the test and thus have no special incentive to study.

In the context of test preparation, there are at least two ways to interpret "fairness." As noted above, one interpretation is that those who have the benefit of being coached may outperform those who don't have someone to mentor their progress, whether or not they have access to prep materials. Another interpretation is that test-prep services might raise the *scores* of their clients without actually improving their readiness for graduate study, and in so doing, deceive graduate programs about these students' likelihood of success.

Unfortunately, there doesn't seem to be any research on whether test-preparation services actually improve the performance of students who go on to graduate school. This would in any event be a difficult experiment to undertake. Many factors influence student success in graduate programs, so even if a student's GRE score improves significantly but their performance remains stagnant, it would be difficult to attribute this solely to test preparation. Other factors confound, including the fact that motivation correlates with both test preparation and success in a graduate program.

So, if we seek an answer to the question of socioeconomic bias, perhaps the best answer is that more-affulent students may indeed perform better, but isn't that the case in almost any educational setting?

## 3. Predictive Validity

A good academic program will endeavor to admit students who will be able to complete the program successfully and graduate. This is in the best interests of both the student and the program. Students who drop out have wasted a significant portion of their life and have suffered a blow to their self-esteem [32] that may be difficult to recover from. Programs that have many failing students in their midst are wasting resources and diverting staff time away from students who would benefit from that interaction. So standardized tests, if they are to be useful, must aid in predicting which students will be able to succeed.

The degree to which a test result may reliably forecast a future event, such academic achievement or professional success, is known as *predictive validity*. Many studies have examined the degree to which the GRE is able to predict various outcomes, such as first-year GPA, cumulative GPA, number of papers authored (research productivity) , and successful degree completion. The results of these studies are not consistent or conclusive, depending as they do on the type and level of the program, the discipline, the sample size and diversity, and the statistical methods and controls.

However, some general observations can be made based on the literature. Kuncel et al. [33] performed a meta-analysis of nearly 100 studies involving 10,000 students. They report operational validities of 0.38 and 0.30, respectively, for the predictive validity of the GRE-V and GRE-Q for masters-level GPAs. For doctoral students, the corresponding values are 0.27 and 0.28. They conclude, "The primary implication of these findings is that the evidence suggests that both doctoral and master's programs can continue to use the GRE and expect that it will provide useful predictive information about their students."

Now, it is not always clear what constitutes a "good" operational validity. An operational validity is a correlation coefficient, so it can range from –1 to 1. A positive score means that higher predictor scores (GRE scores) are associated with higher "criterion scores" (in this case, GPAs). Whether a particular operational validity is "good" depends on the context. If the operational validity is 1.0, that indicates an extremely strong relationship. An operational validity near 0 implies a negligible relationship. In the context of educational testing, an operational validity of 0.3 may be considered acceptable, though in other fields, such as personnel selection, values in the range of 0.5 are sought [34]. Operational validities of 0.3 have also been observed in correlating SAT scores with undergraduate grades [35]. One critique of the GRE is that undergraduate grades are better at predicting graduate achievement [11]. But that does not mean it is useless, because, for example, the combination of SAT + high-school grades have been found to be a better predictor of undergraduate grades than either is alone [35].

Taking a different perspective on predictive validity, a recent meta-analysis by Feldon et al. [36] of 201 studies found that GRE score predicted only 4% of variance in overall GPA. They frame their results from the perspective of *critical whiteness*, which "recognizes that racism is both a historic and a current reality, permeating every aspect of society and experience." They say that education, among other facets of society, has been designed to reinforce systems of power that benefit white communities. They hypothesize that the lower correlations they observe compared to Kuncel et. al. is because the predictive validity of the GRE has diminished as the graduate population becomes more demographically diverse.

In a 2019 study of admissions to 27 Physics Ph.D. programs of diverse national rankings, Miller et al. [37] concluded that neither the GRE Verbal nor the GRE Physics scores had a significant relationship with completion of the program. However, Weissman [38] takes issue with this conclusion, noting that students with low GRE scores who were nonetheless accepted by

competitive programs probably had other factors strongly in their favor, such as letters of recommendation or previous research experience. This tends to make it look like GRE scores do not matter, while the GRE nonetheless helped identify other promising students who subsequently completed their program.

In fact, a problem known to statisticians as *range restriction* makes predicting outcomes based on GRE scores—or any other input—considerably harder than it seems at first glance. The problem is that one knows the outcomes only *for the candidates who were accepted and matriculated*, and they are far from a representative sample, because they impressed the admissions committee enough to be accepted. These are the candidates who are expected to succeed. Their GRE scores likely fall in a much narrower ("restricted") range than all of the candidates who applied. Knowing whether *their* GRE scores are correlated with success does not answer the question about whether the GRE is predictive.

To make this clear, consider an average college class. Does a student's score on the first midterm predict their score on the final exam? Well, someone who gets, say a D, on the first test would be *expected* also to score poorly on the final exam. Some such students would overcome obstacles and improve their performance, but those are presumably the exceptions. Conversely, those who get As on the first exam will very likely get As on the final. So we would probably find that the score on the first midterm is moderately to highly predictive of the corresponding score on the final exam.

But suppose instead, we restrict ourselves to consider only the students who made an A on the first test. Does their relative score on the first test predict where they will fall compared to others on the final exam? Probably we would not be so successful in predicting. Someone who got an A on the first test might well get an A on the final, but we should not expect the top two or three students on the first test to also be the top two or three on the final. The small differences between their performances on Test 1 are too modest to give a firm indication of where they will rank on the final exam.

This same issue arises when we attempt to predict outcomes based on GRE scores. If we *could know* whether the low-GRE students who were rejected *would have* finished their program and graduated, only then could we tell if GRE scores truly predict the outcome of graduation. But we don't know that; we know only the outcomes for the matriculated students. And they may be too close in GRE and/or other characteristics to meaningfully predict differences in outcomes. If we were unaware of range restriction, we might conclude that the GRE does not predict outcomes when actually it does.

There exist statistical techniques to correct for range restriction, and they were employed in the Kuncel meta-analysis [33]. A recent study by Dahlke et al. [39] analyzed SAT results, correcting for range restriction as well as the different courses taken by students. They concluded that with

these adjustments, the SAT (not the GRE!) predicted undergraduate academic performance of women and racial minorities as well as it predicted for white students and men.

Kuncel's meta-analysis found that the GRE was an equally good predictor for masters and Ph.D. programs, but some recent studies call that into question. Petersen et al. [40] looked at STEM programs at four flagship state institutions in the Northeast, following 1805 US citizen and permanent-resident students. They found that for women, the GRE scores of those who completed their program were similar to the scores of those who left, and, counter-intuitively, the scores of men who left the program were *higher* than those of men who completed their program. And Sealy [41] performed a small study using admission data from Vanderbilt University's Initiative for Maximizing Student Diversity. This study collected GRE scores but did not use them in making admission decisions. It involved students from the 1st to 91st percentile on the GRE. GRE scores had no relationship to various metrics of productivity, including publications, time to degree, and faculty evaluations. Although this study was small, consisting of 32 students, 28 of whom had completed their degrees at the time of publication, it represents an attempt to deal with the challenge of range restriction in a field experiment, rather than just with statistical techniques. Feldon [36] mentions this study as raising the possibility that range restriction may have *overstated* the predictive value of the GRE.

## 4. The Holistic Approach

The alternative to using test scores to determine admission is frequently called the *holistic* approach. In a holistic approach to admissions [42], applicants are treated as whole individuals, with several attributes considered, rather than just numerical measures such as grades and test scores. These factors may include academic achievements, such as class rank and rigor of coursework. Extracurricular activities are also included, from the standpoint of participation, but also especially of leadership. Personal statements provide insight into the applicant's goals, character, and aspirations. Letters of recommendation give the perspective of the candidates teachers and mentors. A final factor may be the applicant's demonstrated enthusiasm for the school.

However, holistic review is expensive. Instead of comparisons that can be performed by a computer, it requires considerable staff time to collect and assess all the elements. It may not be affordable for programs in high demand; Harvard, for example, received nearly 6,000 applications for graduate engineering graduate programs last year [43] and the University of Michigan reported over 13,000 [44].

Bastedo et al. [42] performed a mixed-methods study including surveys, focus groups, and a simulation to see how admissions officers implemented holistic review. They surveyed 311 admissions officers at 174 colleges and universities. Their simulation found that holistic review was significantly more likely to admit low-SES (socio-economic status) students. However, holistic review means different things to different people. Highly selective colleges implement it

in a different way than less-selective institutions [45]. The authors observe that confusion about how reviews are conducted seems to induce high-income parents to engage admissions consultants to help their children gain acceptance at their institutions of choice.

Echoing this concern, Woo et al. [4] question whether the criteria espoused by holistic assessment are less biased against low-SES students than test scores would be. Undergraduate

**Table 1.** Summary of arguments for and against use of the GRE

| Pro-GRE | Anti-GRE |
|---|---|
| **Section 2. Objectivity of comparison** | |
| Provides a meritocratic selection process; results not influenced by bias or personal ties [6] | *Questions and scoring may be culturally biased, giving an advantage to European-derived (Western) cultures and a disadvantage to other races* [8]. |
| *Students who engage in self-study for the test do almost as well as students who take a preparation course* [30]. | Socioeconomic bias may favor students who have the money for test preparation [10]. |
| | The cost of the test may discourage less wealthy students from applying to schools that require it [22]. |
| | Non-native speakers of English may be disadvantaged, especially on the verbal test [13]. |
| Since all are aware of how assessment works, results are transparent [7]. | *The administration process may be unfair to students who suffer from test anxiety* [17]. |
| | Tests are not designed to measure soft skills, which are important to success in graduate school [15]. |
| Results can be used as a benchmark to compare performance to national norms [7]. | The GRE may provide an unfair advantage to students who come from selective undergraduate schools [21]. |
| **Section 3. Predictive validity** | |
| The GRE by itself is moderately predictive of graduate GPAs, and likely even better if used in conjunction with undergraduate grades [33, 35]. | *The GRE by itself predicted only 4% of the variance in graduate GPA, according to a meta-analysis of 201 studies* [36]. |
| *Students with low GRE scores who are nonetheless accepted have other factors in their favor, which may influence success* [38]. | |
| Because of range restriction, statistical studies tend to *underestimate* the predictive validity of standardized tests [33]. | *In-depth studies of particular graduate programs have found that the GRE lacked predictive validity* [40, 41]. |
| **Section 4. The holistic approach** | |
| *Holistic review is too expensive for programs that receive large numbers of applications* [43, 44]. | Holistic review considers all aspects of the individual, not just metrics that can be reduced to numbers [42]. |
| *Students from wealthier families may also have an advantage on factors used in holistic review* [4]. | Holistic review is more likely to admit low-SES students [42]. |
| *Lack of transparency in holistic review may encourage well-off parents to use consultants to help their children get admitted* [45]. | |
| *Holistic review may raise suspicions of impermissibly using race as a factor in admisstions* [49]. | |

GPA is not more predictive of outcome than GRE scores, and at least one study [46] suggests that attractive females may get higher grades, indicating that grading is influenced by cognitive biases. Personal statements may be coached or edited by others. Wealthier students would tend to receive more help in writing them. Letters of recommendation exhibit poor inter-rater reliability [47], and may describe applicants of different races or genders differently [48]. The subjectivity of holistic review may invite suspicion of racial bias, especially in light of *SFFA v. Harvard* and *UNC*, so colleges have been urged to at least collect standardized-test scores from students, whether or not they are used in admissions, to defend themselves against charges of discrimination [49].

However, even collecting GRE scores may discourage low-income minority students from applying. First, there is the cost; given the number of test-optional schools, one can avoid the expense by not applying to any that require the GRE [50]. Then there are the large differences in mean scores between races, which may sap the confidence of many minority applicants [4].

## 5. Summary

Table 1 summarizes the arguments presented in the paper. It cites 11 points on each side of the issue. Text in plain Roman presents an argument. If the point on the opposite side of the table is shown in italics, it represents an attempt to answer the original point. For example, the allegation of cultural bias is seen as a retort to the claim that the GRE provides a meritocratic selection process. If a point in italics has no counterpart on the other side of the table, it should be taken as a response to the argument on the line above.

The question of whether to use the GRE is complex. There are many valid points on both sides. Studies that look at outcomes are necessarily dealing with incomplete data, since outcomes cannot be measured for students who are not admitted. Even if we had data for large numbers of student outcomes, that would not put to rest the issues of objectivity and predictive validity. For it is not evident that results for graduate disciplines taken as a whole are valid for any particular discipline, nor that results that pertain to masters programs would also apply to Ph.D. programs.

The main alternative to admitting based upon metrics is holistic review. While it seems to admit more lower-SES students, it would be costly to implement for programs that receive thousands of applications yearly. Because it is less transparent, it may also invite suspicion that admissions are influenced by impermissible factors such as race. Perhaps the best that can be said is that graduate programs should maintain data on admissions factors and procedures and correlate it with outcomes, so that some of the questions posed in this paper can be answered more definitively.

## 6. References

[1]   "History of GRE Test: When, Why and Who Created the GRE Exam," Student Progress. Accessed: Mar. 30, 2024. [Online]. Available: https://www.studentprogress.org/gre/history/

[2]    "Wisconsin State Journal from Madison, Wisconsin," Newspapers.com. Accessed: Mar. 30, 2024. [Online]. Available: https://www.newspapers.com/newspage/397344363/

[3]    "How Does the Old GRE Compare to the Current GRE?," ThoughtCo. Accessed: Mar. 30, 2024. [Online]. Available: https://www.thoughtco.com/old-gre-exam-v-gre-general-test-3211977

[4]    S. E. Woo, J. LeBreton, M. Keith, and L. Tay, "Bias, Fairness, and Validity in Graduate Admissions: A Psychometric Perspective," PsyArXiv, preprint, Aug. 2020. doi: 10.31234/osf.io/w5d7r.

[5]    M. C. Yu and N. R. Kuncel, "The Importance of Standardized Tests in College Admissions," *Counterpoints*, vol. 517, pp. 317–327, 2018.

[6]    G. Hillard, "New Sanity on Standardized Tests," The James G. Martin Center for Academic Renewal. Accessed: Mar. 30, 2024. [Online]. Available: https://www.jamesgmartin.center/2024/03/new-sanity-on-standardized-tests/

[7]    D. G. Meagher, M. Daniel, R. P. Wegner, and N. Ybarra, "Test Bias, Fairness, and Standardized Admission Tests," Pearson PLC, Sep. 2020. [Online]. Available: https://www.aacp.org/sites/default/files/2020-09/test-bias-fairness-and-standardized-admission-tests.pdf

[8]    G. Wiggins, "A True Test: Toward More Authentic and Equitable Assessment," *Phi Delta Kappan*, vol. 92, no. 7, pp. 81–93, Apr. 2011, doi: 10.1177/003172171109200721.

[9]    K. H. Kim and D. Zabelina, "Cultural Bias In Assessment: Can Creativity Assessment Help?," *International Journal of Critical Pedagogy*, vol. 6, no. 2, pp. 129–147, 2015.

[10]   A. W. Hughey, "Why Standardized Testing Is Not Essential in College Admissions," *Counterpoints*, vol. 517, pp. 329–342, 2018.

[11]   "Examining the GRE: Myths, Misuses, and Alternatives - Fairtest." Accessed: Feb. 05, 2024. [Online]. Available: https://fairtest.org/examining-gre-myths-misuses-and-alternatives/

[12]   J. R. and T. Walker, "The Racist Beginnings of Standardized Testing | NEA." Accessed: Feb. 05, 2024. [Online]. Available: https://www.nea.org/nea-today/all-news-articles/racist-beginnings-standardized-testing

[13]   J. E. Kisser, C. R. Wendell, R. J. Spencer, and S. R. Waldstein, "Neuropsychological Performance of Native versus Non-native English Speakers," *Archives of Clinical Neuropsychology*, vol. 27, no. 7, pp. 749–755, Nov. 2012, doi: 10.1093/arclin/acs082.

[14]   K. W. H. Tai and Y. V. Zhao, "Success factors for English as a second language university students' attainment in academic English language proficiency: exploring the roles of secondary school medium-of-instruction, motivation and language learning strategies," *Applied Linguistics Review*, Jul. 2022, doi: 10.1515/applirev-2022-0049.

[15]   P. C. Kyllonen, "Soft Skills for the Workplace," *Change: The Magazine of Higher Learning*, vol. 45, no. 6, pp. 16–23, Nov. 2013, doi: 10.1080/00091383.2013.841516.

[16]   M. Ethier, "GRE Pilots A New 'Soft Skills' MBA Admissions Test At Yale SOM," Poets&Quants. Accessed: Feb. 05, 2024. [Online]. Available: https://poetsandquants.com/2019/03/11/yale-som-pioneers-use-of-soft-skills-mba-admissions-test/

[17]   M. S. Chapell *et al.*, "Test Anxiety and Academic Performance in Undergraduate and Graduate Students," *Journal of Educational Psychology*, vol. 97, no. 2, pp. 268–274, 2005, doi: 10.1037/0022-0663.97.2.268.

[18]   D. E. Powers, "Test Anxiety and the GRE General Test," *ETS Research Report Series*, vol. 1986, no. 2, pp. i–49, 1986, doi: 10.1002/j.2330-8516.1986.tb00200.x.

[19]   D. E. Powers, "Test Anxiety and Test Performance: Comparing Paper-Based and Computer-Adaptive Versions of the Graduate Record Examinations (Gre©) General Test," *Journal of Educational Computing Research*, vol. 24, no. 3, pp. 249–273, Apr. 2001, doi: 10.2190/680W-66CR-QRP7-CL1F.

[20]   J. A. Heissel, E. K. Adam, J. L. Doleac, D. N. Figlio, and J. Meer, "Testing, Stress, and Performance: How Students Respond Physiologically to High-Stakes Testing," *Education Finance and Policy*, vol. 16, no. 2, pp. 183–208, Apr. 2021, doi: 10.1162/edfp_a_00306.

[21]  L. Moneta-Koehler, A. M. Brown, K. A. Petrie, B. J. Evans, and R. Chalkley, "The Limitations of the GRE in Predicting Success in Biomedical Graduate School," *PLOS ONE*, vol. 12, no. 1, p. e0166742, Jan. 2017, doi: 10.1371/journal.pone.0166742.

[22]  "Masters' and doctoral program applications increase as GRE requirements drop," https://www.apa.org. Accessed: Feb. 05, 2024. [Online]. Available: https://www.apa.org/monitor/2024/01/gre-requirements-drop

[23]  A. I. Luppi, C. C. Newton, L. Folsom, E. Galliano, and R. Romero-Garcia, "Ten simple rules for aspiring graduate students," *PLOS Computational Biology*, vol. 17, no. 8, p. e1009276, Aug. 2021, doi: 10.1371/journal.pcbi.1009276.

[24]  P. Montgomery and J. Lilly, "Systematic reviews of the effects of preparatory courses on university entrance examinations in high school-age students," *Int J Soc Welfare*, vol. 21, no. 1, pp. 3–12, Jan. 2012, doi: 10.1111/j.1468-2397.2011.00812.x.

[25]  J. I. Appelrouth, K. M. Zabrucky, and D. Moore, "Preparing students for college admissions tests," *Assessment in Education: Principles, Policy & Practice*, vol. 24, no. 1, pp. 78–95, Jan. 2017, doi: 10.1080/0969594X.2015.1075958.

[26]  R. Moore, E. Sanchez, and M. O. San Pedro, "Investigating Test Prep Impact on Score Gains Using Quasi-Experimental Propensity Score Matching. ACT Working Paper 2018-6.," *ACT, Inc.*, 2018, Accessed: Feb. 06, 2024. [Online]. Available: https://eric.ed.gov/?id=eD593130

[27]  E. Sanchez, "Can Using ACT Online Prep Improve Score Gains? Technical Brief.," *ACT, Inc.*, 2019, Accessed: Feb. 06, 2024. [Online]. Available: https://eric.ed.gov/?id=ED596119

[28]  S. S. Swinton and D. E. Powers, "A study of the effects of special preparation on GRE analytical scores and item types," *Journal of Educational Psychology*, vol. 75, no. 1, pp. 104–115, 1983, doi: 10.1037/0022-0663.75.1.104.

[29]  D. E. Powers, "Understanding the impact of special preparation for admissions tests," *Advancing Human Assessment: The Methodological, Psychological and Policy Contributions of ETS*, pp. 553–564, 2017.

[30]  D. E. Powers and S. S. Swinton, "Effects of self-study for coachable test item types," *Journal of Educational Psychology*, vol. 76, no. 2, pp. 266–278, 1984, doi: 10.1037/0022-0663.76.2.266.

[31]  C. A. Rowland, "The effect of testing versus restudy on retention: A meta-analytic review of the testing effect," *Psychological Bulletin*, vol. 140, no. 6, pp. 1432–1463, 2014, doi: 10.1037/a0037559.

[32]  L. R. Alexitch, "Help Seeking and the Role of Academic Advising in Higher Education," in *Help Seeking in Academic Settings*, Routledge, 2006.

[33]  N. R. Kuncel, S. Wee, L. Serafin, and S. A. Hezlett, "The Validity of the Graduate Record Examination for Master's and Doctoral Programs: A Meta-Analytic Investigation," *Educational and Psychological Measurement*, vol. 70, no. 2, pp. 340–352, Apr. 2010, doi: 10.1177/0013164409344508.

[34]  I.-S. Oh and C. M. Berry, "The five-factor model of personality and managerial performance: validity gains through the use of 360 degree performance ratings," *J Appl Psychol*, vol. 94, no. 6, pp. 1498–1513, Nov. 2009, doi: 10.1037/a0017221.

[35]  K. D. Mattern, B. F. Patterson, E. J. Shaw, J. L. Kobrin, and S. M. Barbuti, "Differential Validity and Prediction of the SAT®. Research Report No. 2008-4," College Board, 2008. Accessed: Feb. 07, 2024. [Online]. Available: https://eric.ed.gov/?id=ED562614

[36]  D. F. Feldon, K. Litson, B. Cahoon, Z. Feng, A. Walker, and C. Tofel-Grehl, "The Predictive Validity of the GRE Across Graduate Outcomes: A Meta-Analysis of Trends Over Time," *The Journal of Higher Education*, vol. 95, no. 1, pp. 120–148, Jan. 2024, doi: 10.1080/00221546.2023.2187177.

[37]  C. W. Miller, B. M. Zwickl, J. R. Posselt, R. T. Silvestrini, and T. Hodapp, "Typical physics Ph.D. admissions criteria limit access to underrepresented groups but fail to predict doctoral completion," *Science Advances*, vol. 5, no. 1, p. eaat7550, Jan. 2019, doi: 10.1126/sciadv.aat7550.

[38] M. B. Weissman, "Do GRE scores help predict getting a physics Ph.D.? A comment on a paper by Miller *et al.*," *Sci. Adv.*, vol. 6, no. 23, p. eaax3787, Jun. 2020, doi: 10.1126/sciadv.aax3787.

[39] J. A. Dahlke, P. R. Sackett, and N. R. Kuncel, "Effects of range restriction and criterion contamination on differential validity of the SAT by race/ethnicity and sex," *Journal of Applied Psychology*, vol. 104, no. 6, pp. 814–831, 2019, doi: 10.1037/apl0000382.

[40] S. L. Petersen, E. S. Erenrich, D. L. Levine, J. Vigoreaux, and K. Gile, "Multi-institutional study of GRE scores as predictors of STEM PhD degree completion: GRE gets a low mark," *PLOS ONE*, vol. 13, no. 10, p. e0206570, Oct. 2018, doi: 10.1371/journal.pone.0206570.

[41] L. Sealy, C. Saunders, J. Blume, and R. Chalkley, "The GRE over the entire range of scores lacks predictive ability for PhD outcomes in the biomedical sciences," *PLOS ONE*, vol. 14, no. 3, p. e0201634, Mar. 2019, doi: 10.1371/journal.pone.0201634.

[42] M. N. Bastedo, N. A. Bowman, K. M. Glasener, and J. L. Kelly, "What are We Talking About When We Talk About Holistic Review? Selective College Admissions and its Effects on Low-SES Students," *The Journal of Higher Education*, vol. 89, no. 5, pp. 782–805, Sep. 2018, doi: 10.1080/00221546.2018.1442633.

[43] "Graduate Student Data | Harvard John A. Paulson School of Engineering and Applied Sciences." Accessed: Mar. 30, 2024. [Online]. Available: https://seas.harvard.edu/prospective-students/prospective-graduate-students/graduate-student-data

[44] "Facts & figures," Michigan Engineering. Accessed: Mar. 30, 2024. [Online]. Available: https://www.engin.umich.edu/about/facts-figures/

[45] A. Mountford-Zimdars and M. Bastedo, "Moving towards more holistic assessment: Selective admissions in the US and England at the brink of the 2020s," in *Routledge Handbook of the Sociology of Higher Education*, Routledge, 2022, pp. 200–212. Accessed: Mar. 30, 2024. [Online]. Available: https://www.taylorfrancis.com/chapters/edit/10.4324/9781003262497-19/moving-towards-holistic-assessment-anna-mountford-zimdars-michael-bastedo

[46] S. N. Talamas, K. I. Mavor, and D. I. Perrett, "Blinded by beauty: Attractiveness bias and accurate perceptions of academic performance," *PloS one*, vol. 11, no. 2, p. e0148284, 2016.

[47] P. M. Muchinsky, "The Use of Reference Reports in Personnel Selection: A Review and Evaluation," *Journal of Occupational Psychology*, vol. 52, no. 4, pp. 287–297, 1979, doi: 10.1111/j.2044-8325.1979.tb00463.x.

[48] J. M. Nicklin and S. G. Roch, "Biases Influencing Recommendation Letter Contents: Physical Attractiveness and Gender1," *Journal of Applied Social Psychology*, vol. 38, no. 12, pp. 3053–3074, 2008, doi: 10.1111/j.1559-1816.2008.00425.x.

[49] S. Porter, "Mandate Standardized Testing," The James G. Martin Center for Academic Renewal. Accessed: Feb. 04, 2024. [Online]. Available: https://www.jamesgmartin.center/2023/12/mandate-standardized-testing/

[50] eSchool M. Contributors, "4 reasons why eliminating the GRE might benefit students," eCampus News. Accessed: Mar. 30, 2024. [Online]. Available: https://www.ecampusnews.com/teaching-learning/2021/03/01/4-reasons-why-eliminating-the-gre-might-benefit-students/