17

Assessment for Admission, Placement, and Outcomes in Higher Education

Wayne J. Camara
Law School Admission Council

Ou Lydia Liu
ETS

Krista Mattern
NBME

Camara, W. J., Liu, O. L., & Mattern. K. (2025). Assessment for admission, placement, and outcomes in higher education. In L. L. Cook & M. J. Pitoniak (Eds.), *Educational measurement* (5th ed., pp. 1189–1264). Oxford University Press.
DOI: 10.1093/oso/9780197654965.003.0017

This chapter focuses on three applications of standardized tests in U.S. higher education—use in admissions, placement, and outcomes assessment. Testing often differs across these three applications, but also within each application across institutions based on institutional need and philosophy. In addition, we address the changes to testing and testing policies that have emerged in the past few years related to the pandemic and other externalities.

Admissions tests serve multiple purposes, some of which are explicitly identified by the testing organization, others that may be proposed by the test sponsors or users (e.g., institution, government agencies), and still others that may be stipulated by policy makers and secondary users of test results. Four purposes or claims have been put forth by ACT concerning the use of admissions test scores: (a) admissions (scholarship), (b) placement (course, program), (c) preparation and academic advising (college readiness, academic planning, application to college), and (d) accountability (school, district, and state). College Board lists five similar purposes for SAT scores: (a) students' college/career readiness, (b) admissions and college course placement, (c) student progress, (d) high school placement, and (e) scholarship decisions (Camara et al., 2019). Graduate and professional admissions tests are much more likely to be used only in admissions and scholarship decisions. Admissions test scores have also been used by policy makers and other stakeholders for purposes that are not intended or sanctioned by test developers, such as for preemployment screening, high school graduation, teacher effectiveness, institutional quality, and ranking. In response, the National Council on Measurement in Education (2019) issued relevant cautionary guidance to dissuade such uses. (See also Ho & Polikoff, this volume, for a discussion of the use of admissions tests for nonintended purposes.)

Placement tests have been described as assessments contributing data to institutional and student decisions about course enrollment, course exemption, and credit by examination for students already admitted to a program or institution (Whitney, 1989). Placement tests are used by many colleges and universities, even those institutions with open admissions policies that do not require testing for admissions, to determine which courses a student should initially complete. They include commercial tests and locally developed tests at individual campuses that are most frequently used for placement in math, reading, and writing/composition. Faculty at many institutions have developed placement tests to determine placement in credit versus developmental courses, as well as the appropriate level in sequential coursework such as foreign languages and mathematics.

Assessments of *student learning outcomes* (SLO) are used by higher education institutions for a variety of purposes, including fulfilling regional and program accreditation needs, internal improvement, faculty/staff interest, equity concerns, and responding to accountability calls (Jankowski et al., 2018). The assessment results are typically examined at the group level without significant direct consequences for individual students. A report by the Commission on the Future of Higher Education released under then–Secretary of Education Margaret Spellings increased the demand and prominence of

SLO assessments. The commission's first report (U.S. Department of Education, 2006) noted a lack of mechanisms for the evaluation of outcomes at U.S. institutions, calling for direct and comparable evidence of learning for improved transparency. The most popular assessments include national student surveys, alumni surveys, locally developed surveys, standardized assessments of general knowledge and skills, rubrics, employer surveys, performance assessments, and portfolios.

CHANGES IN ADMISSIONS AND CURRENT CHALLENGES TO ADMISSIONS TESTING

Admissions tests have been shrouded in controversy since the introduction of the SAT (then known as the Scholastic Aptitude Test) in 1926. The debate about innate intelligence versus achievement, large differences in scores between groups (gender, ethnicity/race), and the influence of coaching on test scores are examples of controversies across the decades (Camara, 2009; Donlon, 1984; Jencks & Phillips, 1998; Lemann, 1999; Willingham & Cole, 1997). However, in 2020, a convergence of externalities emerged, which not only disrupted admissions and admissions testing but also appears to have resulted in profound changes to these processes for the foreseeable future.

COVID-19 and the Social Justice Movement

The coronavirus disease (COVID-19), a severe acute respiratory disease, spread rapidly across much of the globe and was responsible for closing U.S. schools as early as March, 2020. It had a profound impact on admissions and admissions testing for the entering class of fall 2021 and beyond. As of April 3, 2020, all schools in 43 states were closed and most schools in other states were similarly closed (*Education Week*, 2020). Given this closure of schools and colleges, administrations of both the SAT and the ACT were canceled, and initial attempts to administer undergraduate admissions tests resulted in significant reductions in capacity because many test centers remained closed and others introduced social distancing protocols well through 2021. The ACT and SAT tested about 50% of students who registered for testing in the summer and fall of 2020 (Jaschik, 2020). Testing companies sought innovative solutions to bring the admissions tests to students in the fall of 2020; the GRE, Graduate Management Admission Test (GMAT), and Law School Admission Test (LSAT) introduced remote proctored examinations by spring/summer 2020, providing access to thousands of students who tested at home. All three graduate admissions testing programs have continued to employ remote proctored exams for testing at home, while the GRE, GMAT, and, most recently, LSAT have also provided candidates a choice of test center administrations (Camara & Mattern, 2022).

The pandemic coincided with the Black Lives Matter movement (Black Lives Matter: About, 2024), and there was a renewed focus and commitment to equity and social justice by colleges, businesses, and many Americans. A survey of 300 top colleges found that the Black Lives Matter movement influenced changes to admissions processes in

about 30% of institutions (Kaplan, 2021). (See also Cook & Pitoniak, this volume, for a discussion of the impact of the Black Lives Matter movement on assessment.)

Before COVID-19 and the increased spotlight on issues of social justice, about 12% of all 4-year colleges or universities neither recommended nor required admissions tests, which represented about 1% of all freshmen in the United States,[1] with one college test blind (i.e., it did not consider test scores for admissions with any applicants; Camara, 2020). For 2024 admission, over 1,900, four-year colleges (85%) did not require the ACT or SAT for admission and 80 colleges were test blind (FairTest, 2024). Graduate admissions and professional admissions tests were also forced to cancel normal administrations of their tests through spring, resulting in a similar movement to test-optional policies among a number of programs. See the following section for a full discussion of test-optional policies. (See also Ercikan & Flores, this volume, for discussion of the impact of the COVID-19 pandemic on assessment.)

Prior to the pandemic, in early 2019, the University of California's (UC) Academic Senate established an 18-person faculty task force to conduct a study of the university's testing practices, focusing on their role in disproportionate disparities in race, ethnicity, and socioeconomic status (SES) evident in the undergraduate population. This was the third time in less than 20 years that the university system had undertaken a review of its admissions policies and practices to ensure fairness and diversity. The task force was charged with addressing the following questions:

- How well do the UC's current standardized testing practices assess entering high school students for UC readiness?
- How well do the UC's current standardized testing practices predict student success in the context of our holistic, comprehensive review process?
- Should UC's standardized testing practices be improved, changed, or eliminated?
- Do standardized tests fairly promote diversity and opportunity for students applying to UC?
- Does UC's use of standardized tests enhance or detract from UC academic eligibility for high school students?

After conducting meetings, inviting relevant experts in admissions and assessment to make presentations, and completing a comprehensive review of research, the task force released its report with six recommendations (Standardized Testing Task Force, 2020). The recommendations generally encouraged UC to consider additional quantitative factors that could supplement high school grade point average (HSGPA) and tests for eligibility, expand the pool of students eligible for admissions based on class rank, expand academic supports for at-risk students, conduct additional research, and explore the development of a new assessment system.

The task force rejected making admissions tests optional, stating, "With some confidence, we can predict that the following would happen if UC stopped using admissions tests and relied solely on (high school) grade point average (GPA) and other aspects of

the student's transcript for the academic appraisal of applicants" (Standardized Testing Task Force, 2020, p. 85):

- The average student admitted to UC would have a lower first-semester grade point average (FGPA), probability of persistence, probability of graduation, and GPA upon graduation.
- The average financial subsidy by the state would have to rise due to the longer time to graduate per student, and costs per bachelor's graduate would rise.
- Comparisons of in-state and out-of-state applicants would be difficult and unconvincing.
- UC would have less ability to target high-risk students for academic support without tests.

The UC task force similarly dismissed a proposal to adopt the Smarter Balanced Assessment Consortium assessment program in lieu of the ACT/SAT, reasoning that test-taker behavior would change if Smarter Balanced were suddenly used for high-stakes admissions decisions and concluding that diversity-related benefits would not be realized. Camara (2023) compared scores for Smarter Balanced and ACT by ethnicity and race and found that, generally, differences were equivalent, meaning use of Smarter Balanced in place of the ACT would not reduce score gaps for underrepresented minorities. The task force (Standardized Testing Task Force, 2020) found little evidence of instructional validity for achievement tests in general and for Smarter Balanced specifically, noting that since the introduction of Smarter Balanced in California, test performance had been flat, suggesting that additional instruction on the content covered by the assessment shows no increases or effect.

Shortly thereafter, UC president Janet Napolitano made several recommendations to the Regents, which they approved:

1. For fall 2021 and 2022 applicants, campuses will have the option to use ACT/SAT scores in selection consideration if students choose to submit them.
2. Beginning with fall 2023 and ending with fall 2024 applicants, campuses will not consider test scores for admissions selection at all and will practice test-blind admissions selection (California students).
3. During the full period of suspension, from 2021 to 2024, students will have the option to submit an ACT/SAT score for use in scholarship consideration and post enrollment course placement, but not admissions.
4. By 1/21, UC will complete a process to identify or create a new test that aligns with the content UC expects students should have mastered to demonstrate college readiness for California freshmen.
5. If UC is unable to either modify or create a test that meets these criteria and can be available for applicants for fall 2025, UC will eliminate the use of the ACT/SAT for freshman admissions (Office of the President of University of California, 2020).

Napolitano's recommendations were clearly at odds with the task force report, but her recommendations also appeared to consider additional threats imposed by other external factors such as COVID-19, the lawsuit brought against UC,[2] and the persistent and unrelenting challenges to standardized testing at UC. Generally, the UC task force report (Standardized Testing Task Force, 2020) documented evidence that test scores are the single best predictor of academic success at UC; predict well for all major subgroups; are better predictors for underrepresented minorities, first-generation, and low-income students; and had minimal impact on different acceptance rates experienced by subgroups (Camara, 2023).

After the UC Regents adopted a test-optional policy for 2021–2022 and a test-blind policy for 2023, several subsequent events hastened UC's abandonment of test scores altogether. First, a California superior court issued a preliminary injunction stopping the UC system from considering SAT or ACT scores for admissions or financial aid decisions for the class of 2021. The ruling was issued as part of a lawsuit challenging the use of SAT and ACT scores as culturally biased and a violation of equal protection guarantees (*Smith v. Regents of University of California*, 2020). This injunction stated that the test policy adopted by the Regents in the summer of 2020 denied applicants with disabilities equal access to the tests. The ruling reasoned that applicants submitting scores ("submitters") have an inherent benefit denied to nonsubmitters and cited the difficulties of students, particularly those with disabilities requiring accommodations, in registering and completing the ACT and SAT during the pandemic. Next, on May 14, 2021, UC agreed to a settlement with the plaintiffs that made it the first public university system in the nation to stop any consideration of test scores in admissions, scholarship, and financial aid processes (Rippetoe, 2021).

Test-optional policies were clearly gaining in popularity prior to the impact of COVID-19, social justice movements, and UC's task force report. Additional challenges to admissions testing have been attributed to publicized efforts to cheat on tests and gain advantages in admissions, domestically and internationally, and the continued disparity between the demographics of high school graduates and admitted freshmen at many selective institutions. But taken together, these events formed a near-perfect storm, which further fueled alternative admissions processes without test scores.

Test-Optional Policies

Lucido (2018) documented stated and unstated goals of many test-optional policies, which include genuine concerns about the influence and validity of test scores, the desire to enhance the economic and ethnic diversity of the class, and institutional motivations, which may be less altruistic, such as increasing a college's ranking or selectivity (by raising the reported test scores when lower scoring students are nonsubmitters), increasing applications, and increasing yield (since nonsubmitters are more likely to enroll). FairTest (2024) listed over 1,800 colleges and universities that are test optional or test flexible or de-emphasize tests in admissions. Prior to the pandemic, the vast majority of test-optional institutions were private, liberal arts colleges, including 150

for-profit schools and additional institutions with eligibility requirements (e.g., in state, meeting a minimum HSGPA, and submitting other standardized tests). In fact, between 1987 and 2015, only 15% of selective colleges adopted a test-optional policy (Furuta, 2017).

Some institutions that adopted a test optional policy in 2020 have "permanently" moved to test optional with many selective schools continuing their test-optional policies. However, several highly selective institutions announced they were reinstating test requirements for 2025, including Dartmouth, Yale, Harvard, Brown, Caltech, Massachusetts Institute of Technology, Georgetown, Purdue, and the University of Texas at Austin (Hartocollis & Saul, 2024). These institutions cite a variety of reasons for requiring tests, including their own research which shows test scores are the single best predictor of college success at their institutions and the resulting impact when test scores are missing: (a) overreliance on grades with rampant grade inflation (Sanchez, 2023), (b) reliance on the quality and rigor of high schools, (c) rewarding applicants from high schools that regularly send students to the institution and disadvantaging applicants from relatively unknown high schools, and (d) a heightened importance given to personal statements and letters of recommendation – all of which serve to advantage those applicants who are already advantaged (Camara, 2024). Despite uncertainties regarding whether institutions will realize their goals of increased diversity, whether students admitted without test scores perform similarly to students admitted with test scores, and whether undergraduate admissions testing will partially rebound in a post-COVID-19 world, a large majority of institutions are likely to maintain a test-optional policy for the foreseeable future (Camara & Mattern, 2022; Camara, 2024).

Until recently, test-optional colleges had provided only anecdotal evidence that generally claimed there was little to no difference between submitters and nonsubmitters in terms of grades, graduation, or persistence, but evidence allowing comparisons of the two groups in terms of background, high school grades, and other factors was not included (e.g., Hiss & Franks, 2014).

More recently, a few studies of test-optional colleges have investigated these and other issues across multiple institutions. Prior to the pandemic, research showed that among test-optional institutions, 70%–75% of students submitted scores and nonsubmitters were more likely to include a larger proportion of underrepresented racial/ethnic minority students, first-generation students, and low-income students and attain somewhat lower grades in college (Bennett, 2022). One of the first colleges to adopt a test-optional policy was Bates College in Lewiston, Maine, in 1984. Later, a former admission director and colleagues conducted several studies of test-optional colleges. In their most recent study, Syverson et al. (2018) reported that more than half of the test-optional institutions saw larger increases in applications than matched institutions that required testing. However, many test-optional institutions did not see this gain, and the authors failed to explain their criteria and the process used for matching

institutions. The study claimed that test-optional policies did not impact the academic quality of enrolled students and cited a slight increase in the average test score (about 19 points on the SAT scale) and HSGPA (0.02) after going test-optional, but failed to note that test scores were not available for nonsubmitters, who likely had lower scores. The study did report that nonsubmitters had lower high school grades, earned a lower first-semester grade point average (−0.17 on a 4.0 scale), and were slightly less likely to major in science, technology, engineering, and mathematics (STEM), which appears to undermine claims that test-optional policies do not impact academic quality. In a study of 33 schools with test-optional policies, Hiss and Franks (2014) found only marginal differences favoring students who submitted test scores over nonsubmitters in grades (2.88 vs. 2.83 GPA) and graduation rates (64.5% vs. 63.9%), but submitters were more likely to be STEM majors.

Belasco et al. (2015) studied admissions across 180 selective liberal arts colleges, including 32 that were test optional. They found that nonsubmitters were more likely to be from underrepresented and underserved populations. On average, test-optional institutions saw an increase of 26 SAT score points (on a 1,600-point scale), which enhanced their selectivity, and received about 220 more applications than similar institutions that were not test optional; they did not find a difference in ethnic, racial, or economic diversity of enrolled students. Sweitzer et al. (2018) examined the same issues by calculating propensity scores to match test-optional institutions requiring tests, allowing them to model how a test-optional institution would have differed across various outcomes if they had not adopted this policy. Their findings support those of Belasco et al. (2015), with the biggest benefit of test-optional policies being an increase (about 8 points) in SAT scores from a smaller number of applicants and no significant differences in acceptance rates or the diversity of enrolled students. Bennett (2022) compared nearly 100 private institutions that implemented test-optional policies between 2005–2006 and 2015–2016 (a group of earlier adopters) to more than 100 similar institutions that enacted test-optional policies by December 2019 (a comparison group of later adopters). He found enrollment increases for Pell Grant recipients, underrepresented ethnic minorities, and women at test-optional institutions relative to matched peer institutions, but no evidence of increases in applications or yield. Enrollment increases were modest, favoring test-optional schools by 3.1%–4.2% for Pell Grant recipients, but larger for underrepresented minorities (10.3%–11.9%) and women (6.0%–8.0%). Early adopters of test-optional policies did report moderate increases in applications and admission rates, and these benefits may have dissipated as the number of similar institutions with this policy increased over the years. While research seems to clearly suggest that many of the purported benefits associated with test-optional policies are often not supported with data, these policies do not appear to disadvantage institutions.

Data on applications, admissions, and enrollment between 2018–2019 and 2021–2022 from approximately 60 test-optional colleges, representing more-selective public and private 4-year institutions in the United States, were merged with College Board

assessment data to examine how college-going trends and outcomes were affected by pandemic-related disruptions (Edwards et al., 2023; Howell, 2023). The numbers of applications, offers of admission, and enrolled students increased between 2020 and 2021 for nearly all student groups, but more-selective private institutions enjoyed nearly double the increase (29%) in applications compared to all other types of institutions. The proportional representation of ethnic and racial groups changed marginally, while group differences in applications, admission, and enrollment changed very little between 2018 and 2021 among participating institutions. For example, in 2019, the percentage of Black, Native American, and Hispanic/Latino students was 24.4% compared to 25.4% in 2021. The study also found little difference among ethnic and racial groups in their score-sending behavior; students' test scores relative to the college they applied to was the strongest predictor of the decision to disclose their scores (Howell, 2023). In terms of college outcomes, students who disclosed test scores had stronger first-year outcomes (GPA, credits accumulated, retention) compared to students who had no tests or who withheld test scores.

Similarly, Cruce and Sanchez (2023) examined which students were most likely to withhold their ACT scores from test-optional colleges and the impact on persistence to sophomore year. The research focused on matriculants to 142 colleges that introduced a test-optional policy in 2020 and were located in states with an ACT contract to test all public school juniors. The final sample was 102,745 students who attended one of these colleges and graduated from high school between 2017 and 2021. The two predictors examined were (a) discrepancy between students' ACT score and HSGPA and (b) students' ACT score relative to other students' scores at the college. The authors concluded that students make a strategic decision to submit or withhold their test scores, and to help them make that decision, students evaluate their scores against (a) other academic measures (e.g., HSGPA) and (b) the test scores of their peers. Students whose scores are discrepant or who score lower than their peers are less likely to send scores and more at risk of leaving college. Even after all statistical controls were introduced, students who withheld their scores still had a significantly higher likelihood of leaving college within freshmen year or fall to fall.

Chetty et al. (2023) reported that high school grades do a poor job of predicting success at Ivy Plus colleges. Students with a perfect HSGPA of 4.0 achieve a cumulative GPA in college that is less than 0.1 points higher than a student with a 3.2 HSGPA, and higher test scores are associated with higher college GPAs for students from different socioeconomic backgrounds. In addition, students opting to not submit an SAT/ACT score achieve relatively lower college GPAs when they attend an Ivy Plus college.

High School Grades

One important result of test-optional policies is greater reliance on high school grades and other factors. Camara (2018) compared ACT scores of students with an HSGPA of 4.0 or better in the 30 largest schools in a state ($N > 300$ per school) and

reported large between-school differences in the percentage of such students and their mean ACT scores. Schools ranged from having 2.1% to over 23% of students with a 4.0 HSGPA, and average ACT Composite scores differed by over 7 points on a 1–36 point scale. Research on high school grades has revealed systematic grade inflation since the 1990s, which is increasingly beneficial toward wealthy, White, Asian, and private school students (Camara et al., 2003; Hurwitz & Lee, 2018). Consistent with other studies, the UC Academic Senate Task Force (Standardized Testing Task Force, 2020) reported that grade inflation has reduced the predictive power of HSGPA by a third between 2007 and 2015. Chetty and colleagues (2023) reported high school grades have a much weaker association with college GPA. First-year college GPA of students from less advantaged vs. more advantaged high schools, conditioned on SAT/ACT score, were virtually identical, thus displaying no evidence that students from higher-resourced backgrounds outperform students from lower-resourced backgrounds. In addition, reliance on grades alone often results in greater overprediction of college grades for underrepresented minorities than when grades and test scores are used in combination (Mattern et al., 2008; Zwick, 2013), and substantial subgroup differences advantaging White and Asian American applicants on grades remain.

Varsity Blues

In 2019, federal authorities announced charges against over 50 wealthy and famous parents and college personnel in a criminal conspiracy to influence undergraduate admissions at several highly selective colleges, nicknamed Operation Varsity Blues. Parents allegedly paid an admissions consultant between $200,000 and $6.5 million to guarantee their children's admission through a variety of means, including bribing college coaches to place students on lists of recruited athletes regardless of their ability or experience, submitting fraudulent documentation of learning disabilities to secure extra time on admissions tests, and bribing test-site supervisors to correct answers on their child's admissions test after completion (Richer & Binkley, 2019). While the scandal impacted all aspects of competitive admissions, it provided an opportunity for some critics to cite another reason why admissions tests may distort the admissions process (Boost, 2019). However, many saw the incident as an extreme case of distorting admissions in illegal ways, an example of what we have long known about the role of affluence in life and education. Joyce Smith, chief executive officer of the National Association for College Admission Counseling, noted that the scandal reinforces much of what we know:

> that wealthy people have always enjoyed many legal advantages in the process, such as the ability to pay for tutoring, test preparation, and application coaching, or for the truly wealthy to make sizable donations, fund scholarships or endow buildings or faculty seats. And we know that economic status affects one's educational choices at every stage of life. (Hawkins, 2019, p. 15)

Undergraduate Admissions Practices, Testing, and Legal Challenges

The goals of the admissions process are rarely specified in a manner that facilitates validation. Many examinations of undergraduate admissions have described the goals and desired outcomes as multidimensional as well as disguised (Gaff & Meachum, 2006; Rigol, 2003; Willingham, 1985). Desired outcomes from admissions processes include both individual student outcomes (academic ability, ability to benefit, appreciation of lifelong learning, leadership skills) and institutional outcomes, which may be referred to as "building a class." Institutions need to ensure diversity and balance along with many factors, such as financial (ability to pay, need- vs. merit-based aid), talents and interests (to support activities, clubs, and majors), and balance (among other factors, including demographics, geography, and socioeconomic factors).

Admissions processes can generally be classified as mechanistic or holistic. Prior to the 2003 Supreme Court decisions in two affirmative action cases involving the University of Michigan, many large, public universities relied on mechanistic or formulaic components in their admissions processes. In many instances, a minimum test score and/or HSGPA, or a combination of both,[3] may have been used to reduce the large number of applications to a more manageable ratio (applications:admission slots). Mechanistic approaches and cut scores were efficient tools for institutions with large applicant pools, limited budgets, and the desire to have objective criteria to defend admission decisions to the public. (See Ferrara et al., this volume, for a discussion of cut scores and standard setting.) Conversely, private colleges were more likely to conduct holistic reviews, also known as comprehensive or individual review. The holistic review may differ in how it is implemented, but normally involves the review of all materials, by at least two readers, for each applicant, including the student's essay and letters of recommendation (Laird, 2005; Rigol, 2004). Additional components of an applicant's "whole file" may include student accomplishments, opportunities or context, and experiences (e.g., class rank, test score, extracurricular activities, leadership experience, community service). A study examined the relationships between the content of admission essays, self-reported household income, and SAT scores, quantifying essay content using correlated topic modeling and essay style using linguistic inquiry and word count (Alvero et al., 2021). Based on 240,000 admission essays submitted by 60,000 applicants to the UC, they found that essay content and style had stronger correlations to self-reported household income than did SAT scores and that essays explained much of the variance in SAT scores. Overall, there has been little empirical research examining holistic admissions processes, the consistency of ratings attributed to different factors, or the validity of those factors in predicting academic outcomes.

Moreover, an internal audit of undergraduate admissions practice conducted by UC was harshly critical of improper influences and lack of consistent treatment of applicants, finding that campus staff took advantage of weaknesses in admissions processes to admit 64 students as favors to donors, family, and friends. Several weaknesses were

found with the holistic review process, including inadequately trained and supervised reviewers resulting in sometimes overly harsh or overly lenient ratings. They noted that the Office of the President did not review the campuses' admissions processes to detect and prevent unfair or inconsistent practices. Instead, it has allowed weaknesses to persist for years (University of California, Ethics, Compliance, and Audit Services, 2020). Soon after the Supreme Court decision in *Gratz v. Bollinger* (2003), which ruled that the University of Michigan's system of awarding additional points to underrepresented racial/ethnic minority applicants was unconstitutional, many colleges and universities that had not already done so abandoned highly mechanistic admissions processes, and holistic review processes quickly gained popularity. On the same date, the court supported the University of Michigan's Law School in its narrowly tailored use of race in admissions as part of a compelling interest to promote student diversity (*Grutter v. Bollinger*, 2003). To many, a lesson learned from the dual decisions was that race could be considered as part of a holistic review of applicants, but it should not be applied mechanistically with specified points or weights added for minority status. In addition, race should not be a determinative factor in admissions (Rigol, 2004; Zwick, 2018).

Between 1996 and 2012, eight states banned affirmative action in college admissions, and more colleges and universities abandoned strict mechanistic approaches (e.g., cut scores) to make final admission decisions (Davenport et al., 2018). However, the greatest impact from the University of Michigan admissions cases may be the lack of transparency of admissions processes and a decrease in institutional willingness to share data that could disclose subgroup differences among applicants, enrollees, and graduates.

In 2023, the U.S. Supreme Court reversed decades of precedent, effectively ending race-conscious admission programs at colleges and universities. The court found that admissions systems used by Harvard University and the University of North Carolina at Chapel Hill violate the equal protection clause of the 14th Amendment. Chief Justice John Roberts wrote in the majority opinion that "both programs lack sufficiently focused and measurable objectives warranting the use of race, unavoidably employ race in a negative manner, involve racial stereotyping, and lack meaningful endpoints" (*Students for Fair Admissions v. President and Fellows of Harvard College*, 2023, p. 8). The court did say that the decisions should not prohibit colleges from considering an applicant's discussion of how race affected their life in an essay or other means (*Students for Fair Admission v. President and Fellows of Harvard College*, 2023, and *Students for Fair Admissions v. University of North Carolina*, 2023). In both cases, Asian American students rejected by Harvard University claimed they had been disadvantaged by vaguely defined "personal" ratings that invited implicit bias and discrimination, despite being better qualified on academic measures (test scores, grades, extracurricular activities). The plaintiffs used Harvard's internal admissions data and emails to argue that despite better grades and test scores than applicants from other races, Asian Americans were held to a higher standard and penalized by subjective ratings in the holistic admissions process. Harvard University admitted consideration of race was one factor in their holistic admissions process.

A few days after the court's decision was released, civil rights groups filed a complaint against Harvard University's preferences for legacy applicants with the U.S. Department of Education. The complaint noted that nearly 70% of applicants with family ties to donors or alumni are White and are six times more likely to be admitted than non-legacy applicants. About 28% of Harvard's class of 2019 were legacies, meaning fewer admissions slots were available for non-White applicants, who are far less likely to have family ties to the school (*Lawyers for Civil Rights v. President and Fellows of Harvard College*, 2023).

Prior to these decisions, in 2016, *Fisher v. University of Texas et al.*, the Supreme Court ruled that admissions officials may continue to consider race as one factor among many to ensure a diverse student body.[4] In writing for the majority, Justice Anthony Kennedy quoted from another landmark decision: "Considerable deference is owed to a university in defining those intangible characteristics, like student body diversity, that are central to its identity and educational mission," adding that the courts must provide universities substantial but not total deference in designing admissions policy (*Fisher v. University of Texas et al.*, 2016, p. 19).

One important element of holistic admissions processes that has generally escaped scrutiny is the reliability and consistency of ratings. If admissions decisions are to be considered predictive of future success, higher education institutions should provide evidence that ratings of each factor (e.g., leadership), element (e.g., admission essay, letter of recommendation), and student record have high interrater agreement. In addition, validity evidence should be presented that demonstrates the admission processes are accurately identifying students who achieve the desired outcomes of the institution. Legal challenges have largely been directed toward the admissions process and decision, rather than the admission tests. However, the impact of these challenges to affirmative action and the increased demand for diversity of admitted students has had a significant impact on admissions tests, which are addressed later in this chapter.

UNDERGRADUATE ADMISSIONS TESTS AND PRACTICES

The SAT

The inception of the College Entrance Examination Board can be traced to a meeting of the Association of Colleges and Preparatory Schools of the Middle States and Maryland held in Trenton, New Jersey, on December 2, 1899. At that time, there was little agreement among colleges about the subject matter content and preparations that should be required of applicants; College Board was founded to bring order to the chaos of different entrance requirements across established Eastern colleges (Donlon, 1984; see also Clauser et al., this volume). The goal for the fledgling organization was to bring standardization and uniformity to entrance requirements for colleges, but it also had the effect of forcing secondary schools to move from a system built on local autonomy to more common curricula. College Board, formally approved on November 17,

1900, was primarily established to serve as a conduit between colleges and preparatory schools to establish uniformity in admissions standards and preparation/curriculum in secondary schools. It achieved these goals in 1901 by establishing requirements for nine subject matter examinations in English, French, German, Latin, Greek, history, mathematics, chemistry, and physics and administering exams to 973 students, which were read by 39 readers with prominence in their field.

Initially, only 3 colleges replaced their own local exams with the new subject tests, but subsequently, an additional 35 institutions agreed to accept the examinations as substitutes for their own requirements. A decade after their inception, in 1911, the number of students separately examined by Harvard, Yale, Princeton, Penn, Bryn Mawr, Stevens Institute, and the Massachusetts Institute of Technology still exceeded the number of students taking the board exams. College Board examinations continued to consist of a series of essay tests requiring readers to grade each exam; but influenced by the efficiencies and success of intelligence tests during World War I and beyond, the College Board added a comprehensive test of general aptitude to its program in 1926—the Scholastic Aptitude Test (Donlon, 1984; Schudson, 1972). The original College Board achievement tests were transitioned to a fully multiple-choice format in 1937 and placed on the same 200–800 SAT scale; they were rebranded as the SAT II Subject Tests in 2005 and then discontinued in 2021.

The 1926 version of the SAT bears little resemblance to the current test, but it marked the transition from essays to multiple-choice-type items on admissions tests. It contained nine subtests, seven with verbal content and two with mathematical content; and time limits were more stringent, with 315 questions administered in 97 minutes. In 1930, the SAT was split into verbal aptitude and mathematical aptitude, a structure that has remained largely in place over time, including the 2024 design (College Board, 2017a; Lawrence et al., 2002). Other major developments that shaped admissions testing occurring in the 1940s included the transition to machine scoring of the SAT, which replaced the monumental task of human scoring of test questions and eliminated the many hours of training clerical workers, as well as the equating of scores on every SAT form to ensure comparability for admissions decisions. ETS (formerly Educational Testing Service) was founded in 1947 by three nonprofit organizations: the American Council on Education, the Carnegie Foundation for the Advancement of Teaching, and College Board. ETS was established to assume responsibility for test development, administration, scoring, and research, allowing College Board to continue its role as an association and convener of higher education and secondary schools (Fuess, 1950; Zwick, 2006). The use of the SAT increased dramatically during its first 30 years, with over a half-million students testing annually by the late 1950s and over a million students in 1986.

Changes to the SAT were intermittently introduced in content and question type during its history, but the first major revision of the post–World War II SAT did not come until 1994, when antonyms were dropped from the test, reading passages became longer, calculators were allowed, grid-in items were added in math, and the score scale

was recentered in the summer of 1995 based on a full year of operational scores. During this period, the Scholastic Aptitude Test was renamed the Scholastic Assessment Test and then quickly changed to just the SAT I Reasoning Test. Two additional major revisions of the SAT were introduced in 2005 and 2016, and a digital, multistage, adaptive SAT replaced paper-and-pencil version of the SAT in the United States in 2024.

The major change to the SAT in 2005 largely resulted from 2 years of discussions with the UC and was initiated after a speech by Richard C. Atkinson, then UC president, where he recommended elimination of the test as a UC admissions criterion (Schemo, 2001). The most significant change was the addition of a mandatory writing section composed of multiple-choice items and a single 30-minute essay, resulting in now three separate scores (critical reading, mathematics, and writing), each scored on the 200–800 point scale and resulting in a total score of 600–2400. In addition, analogies were removed as an item type in critical reading, quantitative comparison items were removed in the math section, and additional content was added in advanced math topics to better reflect the curriculum.

The 2016 version of the SAT was introduced in March of that year and represented a major redesign of the previous test in terms of content, design, and the interpretation of scores. The redesign was driven by the adoption of the Common Core State Standards by 43 states and a desire to increase market share through statewide adoption (Dudley, 2016; Felton, 2015). In 2013, just 3 states had adopted the SAT for statewide testing, as opposed to 11 states that were using the ACT. The 2016 version of the SAT returned to issuing two scores—evidence-based reading and writing (ERW; which includes a reading test and a writing and language test) and a math score. The essay, which had been required on the SAT since 2005, became optional. Test takers had 3 hours (plus an additional 50 minutes for the optional essay) to complete the test. The two sections—ERW and math—were reported on the 200–800 SAT scale but required a concordance to compare scores with the prior SAT. Other major changes of the 2016 SAT redesign included the elimination of formula scoring, which imposed a penalty for students' incorrect answers, to rights-only scoring, a simple count of correct items, and the inclusion of items in the math section that prohibit calculator use[5] (College Board, 2017a). In 2021, College Board discontinued the optional essay, except for school-based administration with state specific contracts (College Board, 2024c).

Three stand-alone field trials were conducted to develop new score scales, equate initial forms to future forms, and build a concordance between the 2005 version of the SAT and the 2016 redesigned SAT; each is described in the technical manual (College Board, 2017a). A single group design was used to establish a concordance by selecting students who completed the 2005 version of the SAT in the fall of 2015 and having them complete the 2016 redesigned SAT in December 2015. The concordance demonstrated that scores on the 2016 redesigned SAT were substantially higher, with the difference favoring the 2016 SAT between 70 and 100 points on the total score scale (400–1600 scale) (College Board, 2016). College and Career Readiness Benchmarks were revised in 2016. Using historical data from the 2005 SAT linked to college outcomes, the SAT

scores associated with a 75% probability of earning a C or higher were derived. Those scores were than transformed to the 2016 SAT score scale based on a 2005–2016 SAT concordance table to set the updated SAT benchmarks at 530 and 480 for math and ERW, respectively (College Board, 2017b).

The SAT was primarily a paper-based test but began offering a digital administration to states and districts in 2017 to assist with comparability studies. The SAT fully transitioned to a digital format in March 2024, which was partially motivated from persistent security challenges encountered by maintaining a large volume paper program internationally. The multistage test design organized content into two stages, each composed of a module of test items comprising half of the section's items. After answering an initial (routing) module of items representing a broad mix of easy, medium-difficulty, and hard items, students are routed to either a lower- or higher-difficulty second-stage module based on their performance on items in the initial module. Additional changes to the test design included a reduced overall testing time from 3 hours to 2 hours, significant shortening of reading passages with one question per passage, and more time per test item. The reading section is composed entirely of discrete items and shorter passages instead of item sets and extended-length passages, while the mathematics section retains multiple-choice items and student produced-response items. PSAT assessments transitioned to digital tests in the fall of 2023 (College Board, 2024c).

In 2022, 18,513 students were recruited to take both a digital SAT and the previous paper-based SAT within a month. Data from the SAT straight-line concordance study of the digital SAT and paper SAT were used to established a link between the paper-based version of the SAT and the digital SAT's item response theory (IRT) metric for the Reading and Writing section score, Math section score, and performance categories in eight content domains. Scores were linked using the equipercentile method of linking digital IRT maximum likelihood theta values to the paper-based scale scores (College Board, 2024c; see Moses, this volume, for a discussion of scaling, equating, and linking methods).

The ACT

The ACT (then known as the American College Testing) was first introduced in November 1959 in the Midwest region of the United States and was designed with three purposes in mind—admissions, advisement, and placement. The ACT was designed to compete with the SAT, whose primary mission was selection of the few for highly selective institutions (Lemann, 1999). Lindquist (1959) argued that the primary need was "one of finding a type of test that will not just serve a single well-defined purpose, but . . . one of building a multiple purpose rather than a single purpose test" (p. 105). He further argued that such a test should be used more broadly by public institutions and expanded to serve "as the basis for advisement and placement, not just admissions" (ACT, 2009, p. 9). At the time, the SAT remained a tool of primarily selective private institutions. A number of states operated their own statewide assessment system, and there was an absence of a national admissions test, which could be used for students applying to out-of-state institutions. Lindquist argued for the expansion of admissions

testing across the country to a greater number of colleges and well beyond the purview of just elite colleges on the East Coast and to address the needs currently met by a series of tests sponsored by individual states and institutions. Finally, he argued that an admissions test should be based on achievement in curriculum found most often in secondary schools rather than aptitude or innate ability (ACT, 2009; Lindquist, 1959).

The creation of the ACT resulted from both the transformation across higher education in the 1950s, which saw a 433% increase in college enrollment and dramatic increases in educational expenditures, and frustration in attempts to bring significant changes to the SAT (ACT, 2009). Designed by Lindquist and Ralph Tyler to address these purposes, ACT scores were reported to students as well as to colleges; SAT scores continued to be reported only to colleges (Jacobsen, 2017).

The ACT was originally composed of four sections, mirroring the format and content of the Iowa Tests of Educational Development and standard college curriculum: English, mathematics, social studies, and natural sciences. In its first testing year, 132,963 students took the ACT, and that number jumped to nearly 1 million by 1967–1968. At that time, strong regional differences emerged, with 85% of institutions in the Mountain West and Plains states participating in the ACT program compared to 10% for the SAT; conversely, about 17% and 53% of institutions in the northeast and middle states were affiliated with the ACT and SAT, respectively.

Aided by concordance tables first produced in the early 1990s, which allowed admissions officers and high school counselors to compare and convert to a single scale, most colleges accepted scores from either test (Marco et al., 1992). Concordance tables have been produced jointly by ACT and College Board after major redesigns of the SAT in 1994, 2005, and 2016.[6] Similarity of content and strong statistical relationships are required to support the use of concordances to satisfy different goals, such as determining comparable scores, aligning scales, and predicting a student's unavailable ACT or SAT score from their SAT or ACT score (Dorans et al., 1997). The 2018 ACT–SAT concordance tables represent the last time both tests used matched data to establish score comparisons and includes cautions and guidance to assist users in interpreting scores across tests, although recent changes to both the ACT and SAT seem to require a new concordance. In addition, both testing programs have developed assessments for earlier grades (Grades 8–10), which closely mirror the content and statistical blueprints of the ACT and SAT and are vertically scaled to allow interpretation of growth. In particular, the SAT Suite of Assessments includes the SAT, PSAT/NMSQT, PSAT 10, and PSAT 8/9; ACT assessments include the ACT, PreACT (10th grade), and PreACT 8/9.

In 1989, the natural sciences section was transitioned to the science reasoning section and social studies was changed to a reading section (ACT, 2009). In 2005, ACT was also caught up in the debate initiated by Atkinson and the UC and added an optional 30-minute essay. For the 2005 version, ACT consisted of four sections—English, mathematics, reading, and science, with an optional essay. Section scores, a composite score, and a STEM score are reported on the same 1–36 scale used since its inception. The composite score was the average of English, mathematics, reading, and science;

the STEM score is the average of the mathematics and science scores. In 2005, and later updated in 2013 to reflect a more current reference group, ACT established college-readiness benchmarks by empirically deriving the ACT section score associated with a 50% probability of attaining a B or higher in first-year courses in related subject areas (Allen, 2013; Allen & Sconing, 2005), as well as a more rigorous benchmark for readiness in STEM (Radunzel et al., 2015).

For each ACT section test, the construct that is assessed is considered unidimensional—English, mathematics, reading, and science proficiency. In support of this claim, confirmatory factor analyses based on operational data suggest that a one-factor solution has a good model fit for each of the four subject tests (e.g., Li et al., 2017). In 2021, ACT began to also provide an ACT Superscore, which is calculated by averaging students' best section scores across all test attempts to create a new composite score (Radunzel & Mattern, 2020b). The ACT began offering the test digitally, as well as on paper, for school-based administrations (contracts with states, districts, or schools for census testing during a school-day) in spring 2015 and transitioned all international administrations to digital delivery in fall 2018.

In April 2024, ACT announced it was partnering with Nexus Capital Management LP, a Los Angeles–based private equity firm, and transitioning to a for-profit company. The partnership will fund a nonprofit organization conducting services for education and work success and should provide needed capital to stabilize current offerings and grow new products (Arundel, 2024). In February 2024, ACT began offering a digital administration alongside the paper-administration at selected locations[8] and introduced other changes for online test takers in April, 2025, which included making the science section optional (as is the writing section), fewer items, more time per item, reduced total testing time, and a reduction in the number of answer choices in the math section from 5 to 4. The ACT Composite score is now be comprised of only English, math, and reading (requiring only 125 minutes of testing as compared to 195 minutes based on the prior version), while science scores will be reported separately for those who take it.

These changes, which are collectively labelled ACT Enhancements, were introduced for the online national administrations in April 2025 and will be extended to the paper mode in national administrations and all international administrations in September 2025, and finally school-based administrations (under state and district contracts) in spring 2026 (see Table 17.1). ACT notes that the new format is shorter and allows students more flexibility (ACT, 2024b). It has also been noted that students might have been intimidated by the inclusion of science (Knox, 2024).

A linking study was conducted in June 2024 between the Enhanced ACT design and the tradition design with randomly equivalent groups (N = 7,600), confirming the factor structure, showing small differences in the score distributions and lower reliabilities for section scores. The ACT Composite score, now comprised of three section scores (English, reading, math), is not strictly comparable with the ACT Composite score prior to 2025 (English, reading, math, and science) due to both the impact of removing science items from the construct, differences in student performance in science relative

Table 17.1 ACT and SAT Content, Testing Time, and Item Counts

Testing Element	ACT (Enhanced Format)	SAT
Reading section	Reading: 40 minutes, 27 scored items, 9 pre-test items	Reading and Writing: 64 minutes (two 32 minute modules). 50 scored items, 4 pretest items (25 scored items, 2 pretest items per module)
Language section	English: 35 minutes, 40 scored items, 10 pre-test items	
Math section	Math: 50 minutes, 41 scored items, 4 pre-test items	Math: 70 minutes (two 35-minute modules), 40 scored items, 4 pretest items (20 scored items, 2 pretest items per module)
Calculator	Permitted	Permitted
Science section	Optional Science: 40 minutes, 34 scored items, 6 pre-test items	Science: NA
Writing	Optional: 40 minutes essay	NA
Score scale	Composite: 1–36 Section: Reading, English, and Math Optional Sections: Science 1–36; Essay 4 domains and Total 2–12	Total: 400–1600 Section: Reading + Writing 200–800, Math 200–800. Four subscores reported for skills under each section. Reading & Writing = Information and ideas, Craft and structure, Expression of ideas, and Standard English Conventions. Math = Algebra, Advanced math, Problem-Solving and Data Analysis, and Geometry and Trigonometry.
Total testing time	2 hours, 5 minutes	2 hours, 14 minutes
With optional essay and Science	3 hours, 25 minutes	NA
Mode and Digital Format	Paper and Digital available for national and school-based administrations. Digital only for international administrations. Linear format.	Digital delivery. Multistage adaptive (1:2 design).
Pretesting	Embedded within section	Embedded within section

Note. For the ACT, the composite score on the 1–36 scale represents a superscore that is based on English, reading, and math scores. Enhancements to the test are taking place on the following schedule: April 2025 (digital national administrations); September 2025 (paper national administrations); and Spring 2026 (school-based administrations). Information for the ACT is from ACT *Assessment Technical Manual*, 2024a; and ACT Test *Enhancements and Changes*, 2024b; ACT. For the SAT, the digital version was introduced internationally in March 2023 and domestically in March 2024, replacing the current SAT with a revised framework. Information for the SAT is from *Assessment Framework for the Digital SAT Suite* (Version 3.01), 2024a, College Board; and *Digital SAT Launches Across the Country, Completing the Transition to Digital and Providing a Simpler Testing Experience for Students and Educators*, 2024b, College Board.

to other sections, and differences in rounding to derive the Composite score. ACT notes that most students will have a new Composite score that is within the standard error of measurement for the Composite score reported prior to 2025 (Allen & Cruce, 2025, p. 25).

Undergraduate Institutions and College-Going

In 2019–2020, there were 3,982 degree-granting postsecondary institutions with first-year undergraduates. Sixty-seven percent of institutions were 4-year institutions and 33% were 2-year institutions. Among 4-year institutions, 29% were publicly controlled, 58% were privately controlled not-for-profits, and 13% were privately controlled for-profits—a segment that decreased by over 50% between 2013 and 2020. About 25.7 million students are enrolled in U.S. colleges, with roughly 33% enrolled in 2-year institutions and 67% enrolled in 4-year institutions (U.S. Department of Education, 2020). In 2019–2020, there were about 300 public and private not-for-profit 4-year institutions with open admissions in the United States,[9] which indicates that the vast majority of 4-year colleges have some competitive admissions process that considers prior achievement and experience. However, only 75 of these 1,997 institutions examined admit 25% or fewer applicants. Figure 17.1 illustrates that nearly all (86.2%) four-year colleges admitted at least half of all applicants and over half (57.2%) admitted at least three quarters of all applicants (Center for Education Statistics, 2020).

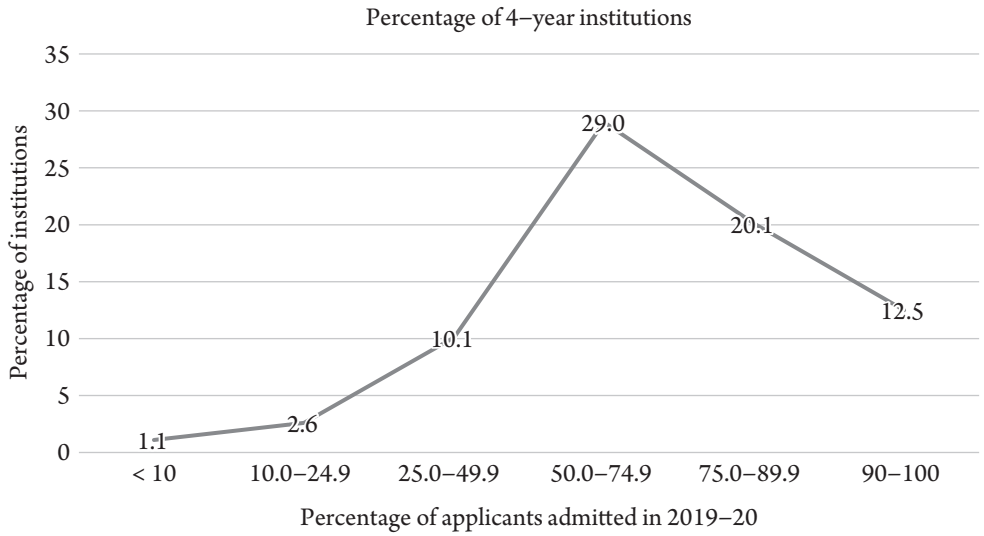

FIGURE 17.1

Percentage of 4-Year Institutions by Undergraduate Admission Rates

Note. The figure does not include the "no application criteria" category, which represents 24.6% of 4-year colleges. Data are from *Digest of Education Statistics, 2020* (NCES 2022-009; Table 219.10), by C. de Brey, A. Zhang, and S. Duffy, 2022, National Center for Education Statistics, Institute of Education Sciences, U.S. Department of Education.

Of the 2.9 million high school completers[10] in 2017, about 67% enrolled in college within 12 months of graduation. Since 1985, 4-year college enrollment has remained fairly stable, at 44% of graduates, while 2-year college enrollments increased by a third, to 23% of high school graduates. Immediate college enrollment rates for students from low-income families was 58%, a rate 50% lower than the enrollment of students from high-income families (Camara & Westrick, 2017). Females continue to outnumber males in immediate college enrollment from high school (72% vs. 61%); and while gaps among racial/ethnic groups have been closing in recent years, discrepancies still existed in enrollment in 2017: 87% of Asian students, 69% of White students, 67% of Hispanic students, and 58% of African American students (U.S. Department of Education, 2019).

An annual survey of admissions trends indicated that applications rose through 2015, with over 80% of students applying to three or more institutions and 35% of students applying to seven or more institutions (Clinedinst, 2019). Despite the media's focus on admissions at the most prestigious universities, only 14% of 4-year institutions accepted less than 50% of applicants in one survey (U.S. Department of Education, 2019), with another survey reporting that figure to be 19% of institutions. While such selective institutions accounted for 37% of all applications, they represented 21% of first-year students (Clinedinst, 2019). Table 17.2 illustrates the prepandemic share of U.S. 4-year institutions, applications, and enrolled students by selectivity.

College admission has remained controversial fodder for the media (Bowen & Bok, 1998; Camara, 2009; Shelton, 1997; Soares, 2012; Zwick, 2002, 2017). Though concerns about the employability of college graduates and their student loan debt have long existed (*The Economist*, 2014), the value of a degree, from both a financial and an individual standpoint, has been well established. Higher levels of educational attainment are associated with reduced unemployment and higher wealth accumulation (Boshara et al., 2015). College degree recipients also engage in more prosocial behaviors, such as volunteering, voting, and participating in political activities (Bowen & Bok, 1998; Goldberg & Smith, 2008).

Table 17.2 Percentage of U.S. Applications and Enrollment by Selectivity at 4-Year Institutions for 2017–2018

Institutional Selectivity	Share of Institutions	Mean Number of Applications per Institution	Share of Applications	Share of Full-Time, First-Time Undergraduates
Accept <50%	19.2	12,492	36.5	21.3
50%–70%	33.4	6,108	31.0	32.3
71%–85%	28.8	5,625	24.6	33.4
More than 85%	18.6	2,771	7.9	13.0

Note. N = 1,599. Adapted from *State of College Admission* (p. 8, Table 2), by M. Clinedinst, 2019, National Association for College Admission Counseling.

The numbers are even more dramatic when recipients of advanced degrees are examined. T. Julian (2012) estimated full-time employees with a bachelor's degree would earn approximately $2.4 million (in 1999 dollars) throughout their adult life compared to $2.8 million for employees holding a master's degree, while those with a doctorate ($3.5 million) and professional degrees ($4.2 million) were estimated to earn significantly more. Given these benefits, it is not surprising that more students choose to attend higher education institutions after high school and that many aspire to earn graduate and professional degrees. Among students completing an admissions test in 2019, approximately 33% aspired to earn a bachelor's degree, and over 43% planned to earn a graduate or professional degree (ACT, 2019a; College Board, 2019a).[11]

Use of the ACT and SAT in Undergraduate Admissions

A wide variety of measures are considered in undergraduate admissions decisions. Grades and strength of curriculum were of considerable or moderate importance for nearly 90% of institutions, followed by admissions test scores at 83% of institutions prior to the pandemic (Clinedinst, 2019). Additional factors such as writing samples, a student's demonstrated interest, letters of recommendation, and extracurricular activities were less important to most institutions. Scores on state assessments were of no importance to over 70% of institutions. As a result of the pandemic and other external factors impacting admissions, as discussed earlier, national data concerning trends in admissions and the importance of various factors have not been updated recently.

Admission test scores provide normative- and criterion-referenced score interpretations, both of which are used in admissions. At institutions where the number of applicants exceeds openings, applicants' performance on admissions factors, including test scores where required or when submitted, are used to sort between students in a typical norm-referenced fashion. Today, test-optional policies reduce reliance on scores to rank order students or establish a minimum score for entry. Previously, students with higher test scores, higher grades, and more rigorous high school coursework were relatively advantaged in the admissions process over students with lower scores, grades, or course rigor (Standardized Testing Task Force, 2020). The issue is not whether a student will go to college, but whether the student will gain admission to more competitive colleges (Beatty et al., 1999). Colleges with higher ratios of applicants to openings may focus more on selecting students into their institutions, but even less-competitive institutions apply normative admissions processes when they must determine which small percentage of applicants is not admitted. Admissions tests may be used in a more criterion-referenced manner if institutions segment applicants into categories such as admit, deny, and further review or if used along with high school grades in a compensatory manner or a sliding scale such as that employed previously for National Collegiate Athletics Association eligibility in Division I sports.[12]

Test Scores and College Search

ACT and SAT scores may be used to advise students, families, and educators prior to and during the transition from high school to college and by colleges to recruit and build their entering class. Test scores can help students and families reduce the number of colleges and universities under consideration from the nearly 4,000 possibilities to a manageable few by identifying potential institutions where they may have a strong academic match (in addition to satisfying other college preferences). One frequent admissions strategy is to identify potential institutions of interest that fit into three categories in terms of probability of admissions (safe, match, and reach colleges; Hoxby & Avery, 2013).

Accountability

Admission test administrations are increasingly funded by states with a variety of goals, such as identifying students who meet empirical benchmarks for college readiness, increasing participation in postsecondary education, and accountability purposes (Camara et al., 2019). In 2018, 26 states funded the administration of the ACT and/or the SAT for all public school 11th graders (i.e., census testing), and half of those states proposed to use scores for federal accountability under the Every Student Succeeds Act (U.S. Department of Education, 2015), which is not without controversy (Marion & Domaleski, 2019; refer to Ho & Polikoff, this volume, for a more detailed discussion). Many states fund either the ACT or the SAT for all juniors, but a few states allow each district to select among admissions tests. State funding of admissions testing began in 2001 with Illinois and Colorado administering the ACT to all juniors. In 2019, about 43% of admissions testing volumes resulted from state funding of admissions testing (Camara et al., 2019; College Board, 2019a). By 2022, 22 states were using an admissions test for high school accountability (Education Commission of the States, 2022).

States view census testing with the ACT or SAT as an opportunity to get their students on the radar of colleges and to get colleges on the radar of all students, especially among underrepresented groups (Anderson, 2014). Benefits associated with using the ACT and/or SAT as compared to a customized state assessments include increased motivation or effort by test takers (Finn, 2015), providing all students with scores accepted for college admissions, increased college enrollment and access (Hyman, 2017; Klasik, 2013), greater security, and reduced testing time (Camara et al., 2019). Several studies have been published that show statewide testing increases overall or 4-year college enrollment in a state, increases access for underrepresented groups, and is a cost-effective intervention (Allen, 2006; Hyman, 2017; Klasik, 2013).

Participation in Admissions Testing

ACT testing volumes increased significantly since the mid 1990's and surpassed the SAT in total U.S. test takers from 2012 until 2017 (Adams, 2017), when the 2016 redesigned SAT regained several state contracts. In 2019, the last year prior

to COVID-19, approximately 1.8 million and 2.2 million high school seniors had taken the ACT and SAT, respectively. Overall, undergraduate admission testing volumes declined by 14% and 22% for the SAT and ACT, respectively, between 2019 and 2023, hitting a low mark for both programs in 2021 and rebounding slightly thereafter. Camara (2024) documented the increase in applicants withholding test scores from 2019–2020 (22.6%) to 2023–2024 (53.2%). Between 2017–2018, over 50% of high school graduates completed the ACT; and between 2018–2020, more than 50% of graduates completed the SAT. Note that totaling percentages from the ACT and SAT may exceed 100% as some students take both tests. In 2021–2022, participation rates fell precipitously with a larger recovery in SAT test takers by 2023.

It is difficult to estimate the number of U.S. high school graduates who took either admission test, both tests, or neither test because test-taker data, understandably, are strictly held by each testing organization. However, the last official concordance between ACT and SAT in 2018 provides some insight. Table 17.3 uses the projected number of U.S. high school graduates and national data reported by ACT and College Board to estimate the percentage of students completing both, only one, or neither test in 2017. Results from the ACT and College Board concordance were based on 589,753 students graduating in 2017 who were matched across both testing programs (College Board & ACT, 2018). For the current analysis, students who took both tests but did not meet the inclusion criteria for the concordance study were retained, resulting in a slightly larger sample size of 593,500, as reported in Table 17.3. In 2017, about 16.5% of graduates took both tests and about 12.5% of graduates took neither test. More recently, both

Table 17.3 Estimated Number and Percentage of U.S. High School Graduates Taking College Admissions Tests in 2017

Admissions Test	Number of High School Graduates	Percentage
SAT only	1,121,981	31.2
ACT only	1,436,538	39.9
Both tests	593,500	16.5
Neither test	450,928	12.5
Total	3,596,650	100

Note. The total number of U.S. high school graduates projected by the National Center for Educational Statistics for 2016 –17, and national data reported by ACT and College Board were used to estimate the percentage of students completing both, only one, or neither test in 2017. The estimate of 593,500 students for the "both tests" value is based on a total of 599,274 students matched across both tests, minus 6,253 international students. The final concordance sample is based on 589,753 students, which excludes students for reasons not relevant to the issue of multiple versus single test participation. In addition, the estimates of students taking both tests are slightly underreported using the concordance samples because they included students who only took the SAT prior to March of their junior year, when a revised SAT was released. Students who took both tests, but one of them was prior to March of their junior year, are counted as taking only one of the tests. Data for projected number of high school graduates are from *Digest of Education Statistics* 2017 (NCES 2018-070; p. 193, Table 219.10) by T. D. Snyder, C. de Brey, and S. A. Dillow, 2019, National Center for Education Statistics, Institute of Education Sciences, U.S. Department of Education. Data for students taking both tests is from *Guide to the 2018 ACT/SAT Concordance*, 2018 (p. 1, para. 3), College Board & ACT.

testing programs have witnessed significant decreases in the number of tested students, coinciding with disruptions in testing due to the pandemic along with a sharp uptake in test-optional admission policies. In particular, the 2021 ACT-tested high school graduating class was comprised of 1.3 million students—which is 375,000 fewer students than the 2020 class (Camara & Mattern, 2022). Similarly, there were nearly 700,000 fewer students in the 2021 SAT-tested high school graduating class, which included 1.5 million students as compared to 2.2 million in 2020. Furthermore, the percentage of college applicants submitting test scores has decreased strikingly from 78% in 2019–2020 to 48% in 2021–2022 (Magouirk et al., 2023).

GRADUATE AND PROFESSIONAL ADMISSIONS TESTS AND PRACTICES

Graduate school and professional degree admissions encompass many academic and professional disciplines (e.g., psychology, economics, English literature, medical school, law school) and a range of degrees (e.g., masters, PhD, MD, JD), housed in large public universities, smaller private and public colleges, and public, private, and not-for-profit stand-alone institutions. This chapter will focus on the admissions processes and tests most widely recognized across competitive academic programs and the largest professional schools—the GRE, GMAT, LSAT, and MCAT. Many additional admissions tests are used for entry to specialty programs in medical fields (e.g., dentistry, optometry, pharmacy, veterinary medicine) but fall outside the scope of the current chapter.

The GRE and GMAT have been administered online and in a computer-adaptive fashion since the 1990s (Graduate Management Admission Council, 2020; Wendler et al., 2014). The MCAT and LSAT transitioned to linear computer-based administrations in 2007 and 2016, respectively (Bingham et al., 2012; Margolis, 2018). In addition to providing an opportunity for improved test security and measurement precision, online testing also allows for the inclusion of more innovative item types and more flexibility in test administration models (e.g., remote proctoring).

Institutions responding to an annual survey of graduate school enrollment and degrees reported approximately 2.1 million applications for admission in fall 2020, with about 68% of applications to master's degree and other (e.g., graduate certificate) programs (Zhou & Gao, 2021). Acceptance rates were 22.3% of doctoral applicants and 55.7% of master's/other applicants, with the largest numbers of doctoral applications in social and behavioral sciences, engineering, and biological and agricultural sciences and those of master's applications in business, mathematics, and computer sciences. Approximately 509,999 first-time graduate students enrolled in fall 2020, with 60.8% female students. Unsurprisingly, given the COVID-19 situation, international graduate first-time enrollment was down by 37.4% between fall 2019 and fall 2020, and a large decline was observed in fields such as engineering, mathematics, computer sciences, and physical and earth sciences (Zhou & Gao, 2021).

Undergraduate admissions processes and tests have received their share of scrutiny in both scholarly research publications and popular media. By comparison, graduate school admissions and tests have been subject to much less examination. A report by the Council of Graduate Schools (Kent & McCarthy, 2016) found that graduate admissions processes are primarily decentralized, with over 75% of academic programs responsible for their own admissions program, in contrast to undergraduate admissions, which are generally centralized within an institution and coordinated by admissions professionals. Graduate admissions decisions are often made by small departments or program-level committees and may include an individual faculty member accepting specific students for doctoral-level programs (Michel et al., 2019). Admissions policies and selection rates differ across disciplines and programs.

In recent years, graduate programs have shown a growing interest in applying holistic review to admission decisions, which suggests that the admissions process includes a broad range of factors (Posselt, 2016). Holistic admissions practices hold great potential in helping programs achieve a targeted balance among admitted student cohorts, particularly in terms of student diversity. Relatedly, there has been a good deal of research suggesting that noncognitive measures provide additional predictive value above cognitive assessments while at the same time increasing the diversity of admitted students (see Shultz & Zedeck, 2011, for research in the law school context). Holistic review is considered a useful, effective, and defensible solution for improving diversity following *Grutter v. Bollinger* (2003); however, decentralized admissions processes present unique challenges with implementation, given limited staff and faculty time. The dominant role played by faculty, the large number of applications, the amount of time required for holistic review, and the long-standing culture have all been cited as reasons that holistic review in graduate admissions often relied heavily on test scores (Kent & McCarthy, 2016).

Admission to medical schools and many law schools remains highly competitive, with the most selective schools exceeding a rate of 20 applications per opening. Test scores and undergraduate grades are typically heavily weighted, among all factors considered in admissions. Many programs compute an index of test scores and grades to aid in ranking candidates, but other factors are also considered in final admissions decisions, including grades in college major, personal statement, letters of recommendation, and, in some instances, personal interviews (Zwick, 2006). Note that personal interviews are common in admissions to medical school but are often optional or not employed in admission to other programs.

Admissions Tests for Graduate and Professional Schools

A persistent criticism of graduate admissions is overreliance on test scores and the use of minimum cut scores to advance in the applicant review process. In a review of admissions processes across 10 highly ranked programs, Posselt (2016) disclosed that the influence of test scores extends far beyond what most departments would

admit or what the testing organizations advise. The Council of Graduate Schools found that minimum test scores are widely used "to eliminate a group of applicants before an initial application review . . . a practice largely rumored but sparsely documented" (p. 8).

A few major trends have occurred in graduate school admissions testing since 2010 that are worthy of mention. First, there is an increased call for holistic reviews, which include cognitive and noncognitive factors in admissions, coupled with a desire to increase the diversity of admitted students (Michel et al., 2019). Second, an increasing number of graduate programs have adopted a test-optional policy. Finally, ETS has actively promoted the use of the GRE as an alternative to the GMAT and LSAT in admissions to business programs and law schools, respectively (ETS, 2017, 2020a). Many business schools and some law schools have moved to accept GRE scores in lieu of GMAT and LSAT scores in an effort to attract more nontraditional applicants (Jaschik, 2018; Olson, 2017). Over 55 U.S. law schools accept the GRE in addition to the LSAT for admission, and more than 90% of business schools are estimated to accept either the GRE or the GMAT for admission. Despite the willingness to accept both scores, there is an absence of research examining institutional preferences and use of scores, score comparability, and comparisons of admission rates and success across testing programs. ETS has produced comparison tools to convert GRE scores to both GMAT and LSAC scores; information on sample size and methodology used in each comparison tool was described by ETS (2017, 2020a; as well as by Klieger et al., 2018). However, the Graduate Management Admission Council (GMAC) and the Law School Admission Council (LSAC) have argued that such score conversion tools are flawed (Han, 2022).

Medical and law school applications represent much smaller cohorts, with approximately 55,000 applicants and 23,000 matriculants to U.S. medical schools in 2023–2024 (AAMC, 2023a, 2023b). First-year enrollment across the 203 American Bar Association–approved schools in 2022 was about 38,000, reflecting a decrease from a peak of more than 50,000 students in 2010. There were 62,545 applicants to law school in 2022, which reflects a decrease of about 25,400 applicants from 2010 (LSAC, 2024).

Tests used for admission to graduate and professional schools have a singular purpose—to predict academic success (grades, graduation). Test scores are also used for selection in fellowship programs, for merit scholarships, and by applicants as a guide in selecting among competitive programs. ETS states that scores from the GRE can be used to identify which applicants are academically prepared for graduate-level study and to measure skills required in graduate and professional schools, including business and law. Inappropriate uses of test scores are generally identified by each testing program and include credit by examination, degree conferral, employment decisions, and use as an outcome measure (ETS, 2022). Recently, a number of law schools have withdrawn from the U.S. News rankings. Table 17.4 provides a comparison of the content, sections, scores, score scale, timing, and number of items across graduate and professional admissions tests.

The GRE

The GRE General Test (hereafter referred to as the GRE) is a cognitive test used for admission to graduate schools in the United States and some other countries (ETS, 2023b). It consists of three measures: verbal reasoning, quantitative reasoning, and analytical writing. Verbal reasoning measures the ability to

> - analyze and draw conclusions from discourse; reason from incomplete data; identify author's assumptions and/or perspective; understand multiple levels of meaning, such as literal, figurative and author's intent
> - select important points; distinguish major from minor or irrelevant points; summarize text; understand the structure of a text
> - understand the meaning of individual words, sentences and entire texts; understand relationships among words and among concepts
>
> (ETS, 2023b, Verbal Reasoning section)

with an emphasis on complex verbal reasoning skills. Quantitative reasoning measures the ability to "understand, interpret and analyze quantitative information, solve problems using mathematical models, and apply the basic concepts of arithmetic, algebra, geometry and data analysis," with an emphasis on quantitative reasoning skills. Analytical writing measures "critical thinking and analytical writing skills, including the ability to articulate and support complex ideas with relevant reasons and examples, and examine claims and accompanying evidence," with an emphasis on analytical writing skills.

Table 17.4 Graduate and Professional Admissions Tests

Test	Sponsor	Test Sections and Scores (Score Scale)	Description of Test Sections	Notes
Graduate Management Admission Test (GMAT) Focus Edition (Total Score =205–805[a])	Graduate Management Admission Council	Data Insights (60–90)	Graphic interpretation, table analysis, multisource reasoning, two-part analysis, integrated reasoning and data sufficiency question types (20 questions, 45 minutes)	The sections are adaptive at the item level. Candidates may select section order. Candidates may bookmark questions in a section for later review; up to 3 answers may be changed. Administered at test centers or at home via remote proctoring.
		Quantitative (60–90)	Problem-solving in applying algebraic and arithmetic knowledge (21 questions, 45 minutes)	
		Verbal (60–90)	This section measures the ability to read and comprehend written material and to reason and evaluate arguments (23 questions, 45 minutes)	

Test	Sponsor	Test Sections and Scores (Score Scale)	Description of Test Sections	Notes
GRE General[b] (Total score = 260–340)	ETS	Analytical Writing (0–6)	A writing assignment to analyze an issue (30 minutes)	Verbal and Quantitative sections are adaptive. Candidates can skip questions or mark them for later review. Administered at test centers or at home via remote proctoring. Candidates can choose which administrations to use in sending scores.
		Verbal Reasoning (130–170)	Ability to analyze and draw conclusions from discourse, reason from incomplete data, understand multiple levels of word meaning, and selecting most distinguishing points, etc. (Section 1: 12 questions, 18 minutes; Section 2: 15 questions, 23 minutes).	
		Quantitative Reasoning (130–170)	Ability to understand, interpret, and analyze quantitative information, problem-solving using math models and basic concepts (Section 1: 12 questions, 21 minutes; Section 2: 15 questions, 26 minutes).	
Law School Admission Test (LSAT)[c] (Total Score = 120–180)	Law School Admission Council	Logical Reasoning (2 Sections)	Ability to examine, analyze, and critically evaluate arguments as they occur in ordinary language (24–26 questions, two 35-minute sections)	Linear digital test. Candidates can skip items and return to them later. The test has been administered at test centers or at home via remote proctoring since August 2023 and will continue through June 2026. A single scale score is reported for the total LSAT.
		Reading Comprehension	Ability to read, with understanding and insight, examples of lengthy and complex materials similar to those commonly encountered in law school (26-28 questions in four sets, 35 minutes)	
		Writing	The test also includes a 50-minute proctored, on-demand writing exam required of all test takers taken remotely and separately from the LSAT. The writing sample is included in the applicant's test record; no score is given.	

(Continued)

Table 17.4 *(continued)*

Test	Sponsor	Test Sections and Scores (Score Scale)	Description of Test Sections	Notes
Medical School Admission Test (MCAT) (Total score = 472–528)	American Association of Medical Colleges	Biological and Biochemical Foundations of Living Systems (118–132)	Solve problems by combining knowledge of biological and biochemical concepts with scientific inquiry and reasoning skills (59 questions, 95 minutes)	Test is 7 hours and 30 minutes long, with breaks. Linear test administered on computers in test centers. Offered on approximately 30 dates.
		Chemical and Physical Foundations of Biological Systems (118–132)	Integrating chemical and physical foundational concepts with scientific inquiry and reasoning skills (59 questions, 95 minutes)	
		Psychological, Social, and Biological Foundations of Behavior (118–132)	Understanding how psychological, social, and biological factors impact health, research methods, and behavior (59 questions, 95 minutes)	
		Critical Analysis and Reasoning Skills (118–132)	Passages and questions measuring ability to comprehend what is read (53 questions, 90 minutes)	

Note. Information for the GMAT Exam is from *Exam Structure, Content, & Features*, 2023; Scores, 2025; and *Understanding Your Score*, 2024b; Graduate Management Admission Council. Information for the GRE General Test is from *The GRE Tests*, 2023c, ETS. Information for the LSAT is from *What to Expect Starting With the August 2024 LSAT*, 2023, Law School Admission Council. Information for the MCAT is from *MCAT Exam Scoring*, 2025a, *What's on the MCAT Exam? Exam Overview*, 2025b; and *What's on the MCAT Exam? PDF Outline*, 2025c, American Association of Medical Colleges.

[a] The previous GMAT score scale, prior to the introduction of the *GMAT Focus Edition* in late 2023, was 200–800 (Graduate Management Admission Council, 2024b).

[b] The GRE design was shortened in September 2023 from 3 hours and 45 minutes to under 2 hours (ETS, 2023c).

[c] The LSAT design changed in August 2024 to eliminate Analogical Reasoning and add a second Logical Reasoning section (Law School Admission Council, 2023b).

In the fall of 2023, the GRE shortened its total testing time to under 2 hours from the 3 hours and 45 minutes it had required. One of the salient features of the GRE has been that for the verbal and quantitative measures, it is adaptive by section with an initial section primarily comprised of questions of medium difficulty; performance on the initial section is used to route test takers into one of three sections of low-, medium-, or high-difficulty questions on the subsequent section. This design allows the GRE to be adaptive at the section level, not at the item level.

The Verbal and Quantitative Reasoning scores are each comprised of two separate sections with a total of 27 items each with total testing time at 41 minutes and 47 minutes, respectively. Previously, the GRE had a total of 80 scored items compared to 54 scored items today. Additional testing time was reduced by removing the unscored

or experimental section used to pretest items and eliminating one of two essay task. Today, an Analytical Writing score is reported from a single 30-minute essay task on analyzing an issue, which requires test takers to evaluate the issue and its complexities and develop an argument with evidence and reasons to support their views (Wendler et al., 2014). GRE test details are also summarized in Table 17.4.

The GRE, then known as the Graduate Record Examinations, were first developed in 1936 as a joint effort in higher education by the graduate school deans of Harvard University, Yale University, Princeton University, and Columbia University and the Carnegie Foundation for the Advancement of Teaching. It was used by the University of Wisconsin in 1938, the first public institution to adopt the GRE. By 1948, over 45,000 students applying to 500 institutions had taken the GRE. The GRE remained a project of the Carnegie Foundation until shortly after the establishment of ETS (then known as Educational Testing Service) in 1947. The GRE was changed from a linear test to an item-level adaptive test in 1993, aiming to improve testing efficiency and the accuracy of estimation of test takers' abilities (Wendler et al., 2014). In 2011, the GRE underwent significant changes (Lewin, 2009), including a new 130–170 scoring scale, the elimination of item types such as antonyms and analogies, the addition of an online calculator for the quantitative section, and the replacement of item-level computer-adaptive testing with section-level adaptive testing (ETS, 2020b). These changes were intended to provide a better test-taking experience and enhance test security. Additionally, the new types of questions in the revised format are intended to test the skills needed in graduate, business, and law school programs (ETS, 2020a). Another feature of the GRE is that it allows test takers to skip questions within a section and change answers within a section. Research suggests that, on average, over 95% of test takers made at least one change to their answers in the verbal reasoning and quantitative reasoning sections; and among those who did change their response(s), 83% were able to improve their verbal reasoning scores and 68% were able to improve their quantitative reasoning scores (Liu et al., 2015).

In addition to its use in general graduate admissions, the GRE was used in admission to more than 1,300 business schools around the world and 91 law schools in the United States as of February 4, 2022. On November 19, 2021, the American Bar Association Council voted to "permit law schools to accept GRE tests scores from applicants in lieu of an LSAT score under Standard 503" (American Bar Association, 2021). Research based on 480 students from 12 institutions shows that both the GRE verbal reasoning and the quantitative reasoning measures were statistically significant predictors of first-year GPA and cumulative GPA for masters of business administration programs (Young et al., 2014). Similarly, research based on 1,587 current and graduated law students from 21 law schools suggests that GRE scores predicted first-year law school GPA when used alone, as well as over and above undergraduate GPA (Klieger et al., 2018). As mentioned earlier, the GRE is administered in Prometric test centers or at home via remote proctoring, which is offered both domestically and internationally with only a few exceptions.

Graduate Management Admission Test

The GMAT is a cognitive test used for admissions into business master and management programs, globally. An updated version of the test labeled the GMAT Focus Edition was released in late 2023[12] (GMAC, 2023). It consists of three 45-minute sections: Quantitative Reasoning, Verbal Reasoning, and Data Insights. The Integrated Reasoning and Analytical Writing sections were removed from the test.

The GMAC, owners of the GMAT, began as an association of 9 business schools in collaboration with ETS in 1953 to develop a standardized test to identify and admit qualified applicants into graduate management programs (GMAC, 2020). The following year, the first GMAT (originally named the Admissions Test for Graduate Business Study) was administered with a testing volume of nearly 3,000 test takers. In 2005, GMAC dissolved its partnership with ETS and selected Pearson and ACT as its new test vendors. In 2018, the number of students taking the GMAT annually peaked at over 200,000, with over 7,000 programs and 2,300 business schools using the GMAT for admission decisions (GMAC, 2018), dropping significantly with the onset of COVID.

The GMAT has gone through various revisions over the years (GMAC, 2020). Notably, the Analytical Writing section was added in 1994. A computer-adaptive version of the GMAT was first introduced to the market in 1997. In 2012, the Integrated Reasoning section was added in response to survey results of nearly 750 management faculty stressing the importance of these skills for management professionals (GMAC, 2011).

In response to COVID-19-related test closures, GMAC began offering an at-home administration option of the GMAT in April 2020. Initially offered as a temporary response, GMAC made this a permanent option and reintroduced test center–based administrations.

GMAC introduced the GMAT Focus Edition in late 2023, composed of three 45-minute multiple-choice sections: Verbal Reasoning, Quantitative Reasoning, and a new section, Data Insights, which replaces the Integrated Reasoning and Analytical Writing Assessment (essay) sections and shortens the test by nearly 1 hour (GMAC, 2023). The Data Insights section presents text, numeric, and graphical information in business scenarios and asks test takers to analyze, interpret, and apply the data. Similar to the GRE, the Quantitative Reasoning and Verbal Reasoning sections of the GMAT exam are computer adaptive; however, for this assessment, the adaptive algorithm works at the item level, selecting items based on the test taker's performance after each item. Three section scores are reported on a 60- to 90-point scale and a GMAT total score is reported on a 205–805 score scale in 10-point increments. Table 17.4 provides additional details on section content and scales. Unlike the previous item-level GMAT CAT, the GMAT Focus Edition allows test takers to mark specific questions within a section and return to them later with the option to change up to three question responses per section (GMAC, 2023).

Law School Admission Test

The LSAT was first administered in 1948 to supplement undergraduate GPA in assisting law schools' admissions professionals (Reese & Cotter, 1994). The

test is owned and operated by the LSAC, a nonprofit corporation in Newtown, Pennsylvania, with more than 200 law schools in the United States and Canada as members. ETS was charged with test development, administration, and scoring through 1979 when LSAC brought these activities in-house. The test's content has been largely unchanged for about 30 years, but two major changes were introduced in July 2019. First, the LSAT transformed to exclusive digital administration utilizing Microsoft Surface Go tablets (Margolis, 2018), with nine national administrations annually. Tablets are stored centrally and distributed to university-based test centers prior to the administration. Second, students taking the LSAT were able to complete the required unscored writing sample in a separate and remote administration. The Digital LSAT consists of four 35-minute sections: two logical reasoning sections, one reading comprehension section, and a variable section used to pretest items that are not operationally scored. Each section is composed solely of multiple-choice items, but students now select the correct option by tapping the screen. A 50-minute required and unscored writing sample is administered digitally and remotely, separate from the test administration.

Law school admissions have fluctuated greatly and the number of LSAC test administrations has also varied over the past 30 years, ranging from 101,000 to over 171,000. Between 1987 and 2019, testing volumes changed an average of nearly 7% annually. Volumes decreased significantly in the late 1990s, increasing in 2002–2011 to a high of over 170,000 and decreasing again to just over 100,000 students in 2014–2017 before rebounding in 2018–2019 (LSAC, 2020). Finally test administrations reached a near-record 169,781 in 2020-21 rebounding from Covid which created barriers to accessing all admissions testing (Kelly & Morgan, 2024). LSAC maintains college transcripts, letters of recommendations, resumes, and personal statements and processes academic credentials for an average of 60,000 law school applicants annually (Anthony et al., 2016). LSAC periodically surveys law school faculty to determine what skills are most essential for success in law school classes; the most recent survey found that higher order critical reading and reasoning skills assessed by the test were ranked highest (LSAC, 2019). Questions used on the LSAT emphasize logic and reasoning skills. Unlike many admissions tests, there is a single scale score on the LSAC (120–180 point scale) computed across 75–76 operational items.

During the pandemic, the LSAT eliminated the variable section and transitioned exclusively to home testing, rebranding the test LSAT-Flex and noting that strict score equivalence with scores from the prior LSAT was not established. In August 2021, the test reintroduced the variable section, retained at-home administration, returned to its original name without the Flex designation, and established score equivalence across versions (Camara & Mattern, 2022). Similar to the GMAC and GRE, in 2023, LSAC introduced dual delivery at home with remote proctoring and test centers during seven 1- or 2-day testing windows. In 2023, the Analogical Reasoning Section was replaced with a second section measuring Logical Reasoning as a result of a settlement to a 2019 lawsuit charging the item type disadvantaged a legally blind test taker (Knox 2023);

and in 2024, a new writing sample was introduced but remained unscored. Table 17.4 provides additional information on the LSAT.

Medical College Admission Test (MCAT)

Introduced in 1928 as the Scholastic Aptitude Test for Medical Students, the admissions test for medical colleges was designed to mitigate attrition rates as high as 50% at some medical schools and included eight separate sections, which were combined into a single score (McGaghie, 2002). Renamed the Medical College Admission Test in 1948, the test had four sections (verbal, quantitative, science, and understanding modern society), each reported on a 200–800 scale. Several other revisions to MCAT occurred subsequently, with the latest major changes introduced in 2015. There was a shift in the focus of constructs from what students know to how well they could apply their knowledge, as well as the introduction of new concepts in biochemistry. In addition, a new behavioral and social science section was added to existing sections, and a new section score scale was introduced (118 to 132) to prevent comparisons to the previous version of the test (Association of American Medical Colleges, 2015). Table 17.4 provides additional information on the four sections, section scale, and total score scale. The MCAT requires more than 6 hours to complete in a single sitting at commercial test centers (Bingham et al., 2012).

A survey of medical school admissions officers found that MCAT total score and undergraduate GPA (total and science/math) were the most important and valued predictors for admission, along with personal interviews at the last stage in the admission process; however, admissions in professional schools now consider many nonacademic factors, suggesting a move to a more holistic approach (Monroe et al., 2013). In 2022, the Association of American Medical Colleges introduced a situational judgment test (named the PREview professional readiness exam) for use by several medical schools in the 2022–2023 admissions cycle. PREview is a standardized exam that presents a series of hypothetical scenarios students may encounter in medical school and asks test takers to evaluate the effectiveness of a series of behavioral responses to each scenario associated with eight competencies (e.g., resilience and adaptability, capacity for improvement, service orientation). It is a remote-proctored, online exam, administered separate from the MCAT (Association of American Medical Colleges, 2022).

RELIABILITY, VALIDITY, AND FAIRNESS OF UNDERGRADUATE AND GRADUATE ADMISSION TESTS

Given the high-stakes nature of admission tests, it is essential that the design and development of these assessments are in adherence with professional standards, producing scores that are highly reliable and have strong validity evidence supporting intended uses (American Educational Research Association [AERA] et al., 2014; see also Huff et al., this volume, for discussion of best practices in test development). Undergraduate and graduate programs use admission test scores to understand applicants' preparedness to succeed in their programs, helping to inform which applicants are admitted.

Given this use, evidence based on relations to other variables is a critical component of the validity argument for admission tests. As such, there is a significant body of literature demonstrating the validity of admissions test scores to predict future educational performance, overall and for student subgroups. This section summarizes the psychometric properties of undergraduate and graduate admissions tests as they pertain to reliability, validity, and fairness.

Reliability

Reliability estimates for undergraduate and graduate admission tests are summarized in Table 17.5. As discussed elsewhere, the ACT, SAT, GRE, and GMAT all underwent significant redesigns between 2023 and 2025, which entailed reducing testing time and the number of items, overall and per sections. In many instances, section score reliabilities were reduced from earlier designs in exchange for other benefits (e.g., less testing time, reduced speededness). Both at the undergraduate and at the graduate level, standardized tests used for admissions are highly reliable, with estimates for the total or composite score typically reaching values of .90 or higher, exceeding the minimum reliability thresholds proposed for high-stakes applications in the past (Nunnally, 1978).

Table 17.5 Reliability of Undergraduate and Graduate Admissions Tests

Test	Test Sections	Reliability
ACT	English	.88
	Math	.88
	Reading	.84
	Science	.85
	Composite	.96
SAT	Evidenced-Based Reading and Writing	.94*
	Math	.90*
	Total	.96*
Graduate Management Admission Test (GMAT)	Quantitative	.90
	Verbal	.89
	Data Insights	.86
	Total	.94
GRE	Analytical Writing	.76
	Verbal Reasoning	.87
	Quantitative Reasoning	.93
Law School Admission Test (LSAT)	Total Test Score	.90

(continued)

Table 17.5 *(continued)*

Test	Test Sections	Reliability
Medical College Admission Test (MCAT)	Biology and Biochemical Foundations of Living Systems	.82–.86
	Chemical and Physical Foundations of Biological Systems	.82–.86
	Psychological, Social, and Biological Foundations of Behavior	.82–.86
	Critical Analysis and Reading Skills	.82–.86
	Total Score	.95

Note. Data for the ACT Test are from *Initial Evidence Supporting Interpretations of Scores From the Enhanced ACT Test* (ACT Research Report, p. 12), by J. Allen and T. Cruce, 2025, ACT. Data for the SAT are based on the version that existed *prior to* the introduction of the digital SAT in March 2023 internationally and in March 2024 domestically, and are in the *SAT Suite of Assessments Technical Manual Appendixes* (p. 70, Table A-6.2), by College Board, 2017b. Reliability coefficient estimates are not reported in the technical manual for the new digital SAT, but the reliability of linear digital scores is comparable to that those for its multistage tests per the *Digital SAT Suite of Assessments: Technical Manual* (p. 48), by College Board, 2024c. Data for the GMAT are from *GMAT Focus Psychometric Updates* [Seminar presentation], by K. T. Han, 2025, October 17, GMAT Test Preparation Summit, Reston, VA, United States. Data for the GRE are an average of the reliability estimates for all multi-stage tests administered between September 2023 and May 2024 to reflect the reliability of the shortened GRE, and are in the *GRE Guide to the Use of Scores* (p. 34, Table 4A), by ETS, 2024. Data for the LSAT are from *2021–22 Interpretive Guide for LSAT Score Users*, by Law School Admission Council, 2022. Data for the MCAT are from *The MCAT Exam: Year at a Glance*, by Association of American Medical Colleges, 2015.

Reliabilities exceeding .90 maybe unrealistic and unnecessary for section scores since total scores are primarily used in admissions decisions. For more detailed information, consult the references listed in the notes section of Table 17.5. (See also Lee & Harris, this volume, for a discussion of reliability.)

Validity

For each of these testing programs, a substantial collection of research evidence exists demonstrating the relationship between admission test scores and subsequent higher education success. This section summarizes this body of literature, with common themes emerging. First, test scores are predictive of future performance across multiple indicators of success (e.g., first-year grades, cumulative GPA, retention, graduation), though the majority of research tends to focus on first-year GPA. This is largely due to the expediency with which first-year grades can be collected as compared to more distal outcomes, coupled with more standardization in course taking in the first year as compared to later years. That said, it is still critical and common practice to investigate the validity of test scores for predicting additional metrics of student success. Second, the research points to the incremental validity of test scores for predicting success, above and beyond other admission criteria, most notably prior grades. In particular, studies have examined the degree to which test scores add to the prediction of first-year GPA above HSGPA in the context of undergraduate admissions and above undergraduate GPA (UGPA) in the context of graduate admissions. Results highlight the unique information provided by test scores, underscoring the value of taking a multiple-measures approach in the admission decision

process. Third and finally, for graduate programs, the research evidence suggests that test scores are more predictive of graduate performance as compared to UGPA, whereas HSGPA tends to be slightly more predictive of college success as compared to test scores in the undergraduate context. Refer to Table 17.6 for a summary of validity evidence across the testing program for a focal outcome of interest: first-year GPA; a deeper review of existing validity evidence for each testing program is provided next. (See also Lane & Marion, this volume, for discussion of validity and validation.) Most of the evidence described is based on older test designs prior to changes introduced after 2023. Similar data on the shortened ACT, GRE, and GMAT will be forthcoming.

ACT

Numerous large-scale, national studies have been conducted linking ACT scores to success in college, including performance in the first year, credit-bearing courses, first-year GPA, retention, progress toward a degree, cumulative GPA, and degree completion (e.g., Allen, 2013; Radunzel & Noble, 2012; Sawyer, 2010, 2013; Westrick et al.,

Table 17.6 Validity of Undergraduate and Graduate Admissions Tests for Predicting First-Year Grade Point Average (GPA)

Test	Test Correlation	Prior GPA Correlation	Multiple Correlation
ACT	.51	.61	.64
SAT	.57	.54	.66
Graduate Management Admission Test (GMAT)	.48	.30	.53
GRE	.34–.38	.33	n/r
Law School Admission Test (LSAT)	.60	.42	.66
Medical College Admission Test (MCAT)	.57	.52	.65

Note. For ACT and SAT, first-year GPA is freshman year in undergraduate programs. For other tests, GPA reflects first year of graduate or professional school study. ACT correlations are for paper-based version of the test, prior to the 2024 -2025 enhancements, and are corrected for indirect range restriction and criterion unreliability values. Data are from "The Road to Retention Passes Through First-Year Academic Performance: A Meta-Analytic Path Analysis of Academic Performance and Persistence," by P. A. Westrick, F. L. Schmidt, H. Le, S. B. Robbins, and J. M. Radunzel, 2021, *Educational Assessment, 26*(1), p. 43, Tables 2 and 3. SAT correlations are corrected for restriction of range. Data are from *Digital SAT Suite of Assessments: Technical Manual* (p. 172, Table 8.12), 2024c, College Board. GMAT correlations are for the version prior to the Focused Edition, for which data are not available at this time. Data for the prior version are from *GMAT Exam User Guide for Institutions* (p. 9, para. 5), by Graduate Management Admission Council, 2018. GRE correlations are for the prior design, and are corrected for range restriction and criterion unreliability. Only bivariate correlations between individual GRE sections and first-year graduate GPA are reported. Multiple correlations were not reported (n/r). For predicting the unit-weighted composite of cumulative graduate GPA and faculty ratings, the multiple correlation for three GRE scores was .45; adding undergraduate GPA increased the multiple correlation to .50. Data are from "A Comprehensive Meta-Analysis of the Predictive Validity of the Graduate Record Examinations: Implications for Graduate Student Selection and Performance," by N. R. Kuncel, S. A. Hezlett, and D. S. Ones, 2001, *Psychological Bulletin, 127*(1), p. 169, Table 2, and p. 175, Table 9. LSAT correlations are corrected for multivariate range restriction. Data are from *Summary of 2017, 2018, and 2019 LSAT Correlation Study Results* (2nd ed., p. 24, Table 3.2b), 2023a, Law School Admission Council. MCAT correlations are corrected for range restriction. Data are from "The Validity of Scores From the New MCAT Exam in Predicting Student Performance: Results From a Multisite Study," 2020, by K. Busche, M. L. Elks, J. T. Hanson, L. Jackson-Williams, R. S. Manuel, W. L. Parsons, D. Wofsy, and K. Yuan, in *Academic Medicine, 95*(3), p. 394, Figure 3.

2015, 2021). For example, Allen (2013) showed that students who earn higher ACT scores are more likely to earn a B or higher in first-year, credit-bearing courses in related content areas. In addition to specific course grades, students with higher ACT scores are more likely to earn higher GPAs in the first year of college (Sawyer, 2010, 2013; Westrick et al., 2015, 2021). In particular, based on nearly 170,000 students attending one of 50 higher education institutions, Westrick et al. (2015) estimated a mean correlation of .51 for ACT Composite score with first-year GPA, after correcting for multivariate restriction of range. In a follow-up study based on the same data set, Westrick and colleagues (2021) demonstrated the unique contribution of ACT scores above and beyond HSGPA in terms of predicting first-year GPA. In particular, they estimated correlations of .51 for ACT, .61 for HSGPA, and .64 for the combination of the two measures for predicting first-year GPA; all reported correlations were corrected for indirect range restriction and criterion unreliability.

A good deal of evidence also exists demonstrating the relationships between ACT scores and more long-term college outcomes, even after controlling for HSGPA (Mattern & Allen, 2016). For example, Radunzel and Noble (2012) showed that students with higher ACT scores were more likely to achieve a higher cumulative GPA (i.e., 3.00, 3.50, and 3.75) and were more likely to earn a college degree within 6 years. Finally, accompanying ACT's announcement that it is exploring additional test features such as superscoring, section retesting, and online testing on national test dates (ACT, 2019b), a handful of validity studies supporting the new test options have been published (Mattern et al., 2018; Mattern & Steedle, 2020; Radunzel & Mattern, 2020a, 2020b). Validity evidence for the ACT Enhancements were not yet available at the time this volume was published.

SAT

Similar to the ACT, countless studies have been conducted evaluating the predictive validity of SAT scores, alone and in combination with HSGPA, in terms of college success (e.g., Berry & Sackett, 2009; Bridgeman et al., 2000; Kobrin et al., 2008; Mattern & Patterson, 2014; Westrick et al., 2019). These studies demonstrate that SAT scores are predictive of both short- and long-term college success, including course grades, first-year GPA, retention, and degree completion, alone and in combination with HSGPA.

With the redesign of the SAT in 2016, College Board has been conducting numerous studies to document the predictive validity of the revised assessment (Marini et al., 2019; Shaw et al., 2016; Westrick et al., 2019, 2020). Based on over 223,000 students attending one of 171 postsecondary institutions, results from the first national validity study of the redesigned SAT found a correlation of .51 for SAT total and .53 for HSGPA for predicting first-year GPA (Westrick et al., 2019). When the two measures were used in combination, the multiple correlation increased to .61, demonstrating the unique predictive power of each measure; reported correlations were corrected for multivariate range restriction. In preparation for the domestic launch of the digital SAT in

March 2024, College Board conducted a concurrent validity study with 11 undergraduate institutions representing nearly 2,000 first-year, first-time students (College Board, 2024c; Marini et al., 2023). Corrected correlations for restriction of range (and uncorrected correlations) between the Digital SAT with first-year GPA were:

.57 (.39) SAT Total Score with FGPA
.53 (.32) SAT Reading and Writing Section Score with FGPA
.55 (.35) SAT Mathematics Section Score with FGPA
.66 (.46) SAT and HSGPA with FGPA
.54 (.27) HSGPA with FGPA
.12 (.19) SAT incremental validity beyond HSGPA with FGPA.

The study indicated that the relationship between the digital SAT and first-semester GPA was comparable to that of the paper-and-pencil version and that the digital SAT adds incrementally to the prediction of college performance above and beyond HSGPA, providing initial validity evidence for the use of the digital SAT for college admissions. Digital SAT Total scores and HSGPA individually have moderately strong, positive relationships with the number of first-year credits earned, with correlations of .45 and .39, respectively, and a multiple correlation of .50 when employed together (College Board, 2024c).

GMAT

Hundreds of studies have been conducted demonstrating the utility of GMAT scores for predicting success in graduate business programs (Kuncel et al., 2007; Oh et al., 2008; Talento-Miller, 2017). For example, based on meta-analytic results including over 400 independent samples and nearly 65,000 students, Kuncel et al. (2007) found that both GMAT scores and UGPA were predictive of graduate school outcomes including first-year graduate GPA, cumulative graduate GPA, and persistence. For the first-year graduate GPA results, they reported correlations of .47 for GMAT Total and 0.35 for UGPA, after applying corrections for direct range restriction and criterion unreliability. They concluded that the combination of GMAT scores and UGPA would likely result in superior predictive power. Based on the same data set, a follow-up study by Oh and colleagues (2008) suggested that the predictive power of the GMAT is even higher, estimating a correlation of .51 between GMAT Total and first-year graduate GPA when applying different restrictions of range assumptions. GMAC (2018) reported similar findings, with a correlation of .48 for GMAT Total and .30 for UGPA, as well as a multiple correlation of .53 when the two measures are used in combination to predict first-year graduate GPA.

More recently, a national validity study including 28 graduate business programs was conducted, and the results aligned with previous findings (Talento-Miller, 2017). In particular, GMAT Total is a strong predictor of graduate GPA (.38) and shows a stronger relationship with graduate performance as compared to UGPA (.32). Moreover, this study provided validity evidence in support of the Integrated Reasoning and

Analytical Writing sections, reporting median correlations with graduate GPA of .27 and .13, respectively. Finally, the combination of the four GMAT section scores and UGPA resulted in the strongest prediction of graduate GPA, with a multiple correlation of .51. Predictive validity for the newly introduced GMAT Focus Edition was not available at this time.

GRE

Many studies have investigated the GRE scores' prediction of important academic outcomes in graduate programs, such as first-semester GPA, first-year GPA, cumulative GPA, progress in the program, comprehensive examination performance, advisor and instructor ratings, research productivity, degree attainment, and time to degree (e.g., Klieger et al., 2014; Kuncel et al., 2001, 2010; Powers, 2004; Powers & Fowles, 2000). For instance, a meta-analysis consisting of 1,753 independent samples of over 80,000 students showed that both GRE verbal reasoning and quantitative reasoning were strong predictors of first-year and cumulative graduate GPA and that the predictive validity of GRE scores was as strong as that of UGPA (Kuncel et al., 2001). Additionally, the study demonstrated the incremental validity of GRE scores above and beyond UGPA in the prediction of cumulative graduate GPA.

Another predictive validity study of the GRE analyzed data from 10 public universities in Florida based on 21,127 master's students and 4,229 doctoral students from 28 disciplinary domains (Klieger et al., 2014). The results showed that GRE analytical writing was as strong a predictor of graduate GPA as verbal reasoning and quantitative reasoning scores. It is unclear how graduate programs use analytical writing scores in admissions, yet research shows that the analytical writing score is a strong predictor of graduate GPA and is also a good indicator of students, likelihood to be in academic peril, defined as having an average C+ or lower GPA (Bridgeman, 2016; Klieger et al., 2014). Several meta-analyses of GRE's predictive validity have reported small to moderate, positive correlations between scores and graduate first-year and cumulative GPA ($.20 \leq r \leq .36$). A more recent meta-analysis concluded that GRE validities have declined slightly over time, with an average correlation between GRE and student outcomes of .22 in 1957 compared to .15 in 2022. Similarly, the average correlation between GRE and GPA decreased from .25 in 1957 to .16 in 2022 (Feldon et al., 2023).

Researchers have also investigated the GRE's use in graduate admission in international contexts. Results based on a leading research university in Singapore revealed that both GRE verbal reasoning and quantitative reasoning scores predicted first-semester and first-year graduate GPA (Liu, Klieger, et al., 2016). They also predicted program standing (students in good standing are those who never received an academic warning, academic probation, or dismissal). For example, the difference in GRE quantitative reasoning score between students in good standing and those not in good standing was as large as .86 standard deviations for master's programs. Another study analyzed a multinational sample of students at

a university in the Netherlands, and the results indicated that the GRE analytical writing score predicts graduate GPA and master's thesis grades above and beyond UGPA (Schwager et al., 2015; Zimmermann et al., 2017). Furthermore, such prediction is independent of students' SES as indicated by the highest parental degree. Additional empirical validation studies on the shortened GRE should be available in the future.

LSAT

Routinely, the LSAC conducts national validity studies examining the relationship between LSAT scores and first-year performance in law school, alone and in combination with UGPA (Anthony et al., 2013, 2016; LSAC, 2023a; Stilwell et al., 2007, 2009, 2011). Based on the data from the most recent national validity study, which included 163 schools that participated in LSAT Correlation Studies in the 2019 cycle, LSAC reported a correlation of .40 for LSAT and .26 for UGPA in terms of predicting first-year law school GPA (2023a, p. 24, Table 3.2a). When combined, the multiple correlation of LSAT scores and UGPA with law school GPA increased to .49. After correcting for restriction of range, the correlation between LSAT and first-year law school GPA increased to .60; a similar jump was seen for UGPA (.42). A corrected multiple correlation between LSAT and UGPA with first-year law school GPA of .66 was reported (p. 24, Table 3.2b). Consistent with previous LSAC validity studies, the results highlight the superior predictive power of LSAT scores as compared to UGPA and indicate that the combination of the two measures results in the best prediction of law school performance.

MCAT

A good deal of research has been conducted demonstrating the value of MCAT scores for predicting important outcomes for medical students, including grades earned in medical school, performance on the United States Medical Licensing Examination, and unimpeded progress toward graduation (Donnon et al., 2007; Dunleavy et al., 2013; E. R. Julian, 2005). Across these various operational definitions of medical school success, results of these studies have consistently found that MCAT Total scores are predictive of medical school performance and that they have larger validity coefficients as compared to the results based on UGPA. Moreover, MCAT Total scores provide incremental validity above and beyond UGPA, with the combination of the two resulting in the best prediction of medical school performance.

Similar to several of the other testing programs, the MCAT went through a major revision, requiring the collection of new validity evidence to support the use of the revised assessment for medical school admissions. In 2013, a concurrent pilot study was conducted with 11 participating institutions, providing initial evidence of the predictive validity of the new Psychological, Social, and Biological Foundations of Behavior section (Association of American Medical Colleges, 2015). The results

indicated that the new section correlated more highly with medical school grades (.41) as compared to scores from the old exam (ranging from .12 to .30). Busche et al. (2020) conducted an initial validity study of the new 2015 MCAT assessment, examining the validity of MCAT scores, alone and in combination with UGPA, in terms of predicting first-year medical school performance and on-time progression to the second year. The study included roughly 1,000 students who took the new MCAT and entered medical school in 2016 at one of the 16 participating institutions. The results found MCAT Total scores to be predictive of first-year medical school GPA, with a median correlation coefficient of .57 as compared to .52 for UGPA. Based on the combination of both measures, the multiple correlation was .65, highlighting the unique predictive power of both measures. Reported correlations were corrected for range restriction.

Fairness

The fairness of standardized tests is often heatedly debated in the popular press and is the focus of much research of admission tests (refer to Zwick and Rodriguez & Thurlow, this volume, for a more in-depth discussion of fairness in educational measurement and score interpretation). The belief that standardized tests are biased tends to revolve around the finding that there are large performance differences by demographic characteristics (e.g., race/ethnicity, SES). This is true for all testing programs (Association of American Medical Colleges, 2015; ACT, 2024; Busche et al., 2020; College Board, 2017a; Dalessandro et al., 2014; ETS, 2020b; Talento-Miller, 2017), as well as for other measures of educational achievement, such as grades and educational attainment level. That said, subgroup differences, in and of themselves, do not establish test bias (AERA et al., 2014). However, when such findings are observed, follow-up analyses should be conducted to determine what may be contributing to these differences. Subgroup differences by ethnicity/race or SES "do not originate with college admissions tests but are instead reflective of long-standing, systemic variation in educational resources and opportunities that also manifest themselves in achievement disparities on standardized tests at the K–12 level" (Bennett, 2022, p. 183).

More globally, admission testing programs tend to follow a similar protocol to ensure the fairness of their tests (e.g., ACT, 2024; College Board, 2017a; GMAC, 2018; Wendler et al., 2014). This includes providing candidates with appropriate testing accommodations for special populations, conducting extensive item and form reviews by independent fairness experts, analyzing items for differential item functioning, and examining test data for subgroup differences, differential validity, and differential prediction. More detailed research on the fairness of admission tests with a focus on differential validity and prediction results is described below for each testing program. Most data are based on older test designs prior to changes introduced after 2023.

ACT

Differential validity and prediction by student subgroups have been the focus of much research on the ACT (e.g., Allen, 2016; Radunzel & Noble; 2013; Sanchez, 2013). For example, Sanchez (2013) examined the degree to which the ACT exhibited differential prediction of first-year GPA by race/ethnicity, gender, and income level. Results showed that the strength of the relationship was similar across subgroups, with slight overprediction for underrepresented racial/ethnic minority students, males, and low-income students. Based on the same student subgroups, Radunzel and Noble (2013) reported similar results when examining the differential prediction of more distal college outcomes, including cumulative GPA and degree completion. A subsequent study by Mattern et al. (2017) found that the underprediction of female performance was reduced roughly by half after controlling for conscientiousness. Contrary to the belief that test scores are biased against underserved students given lower average scores, research consistently shows that underserved students tend to perform worse in college than predicted based on their ACT scores.

SAT

Differential validity and prediction results for the SAT mirror the ACT findings. In particular, the strength of SAT scores for predicting important college outcomes, including first-year GPA, cumulative GPA, and graduation, is similar across student subgroups (Marini et al., 2019; Mattern et al., 2008; Mattern & Patterson, 2014; Young, 2004). SAT scores tend to slightly overpredict college performance for underrepresented racial/ethnic minority students, students from lower SES backgrounds, and males; the extent of overprediction associated with HSGPA exceeds that of SAT scores for underrepresented racial/ethnic minority students. Various explanations for differential prediction have been offered, such as differences in course-taking behavior and noncognitive factors (Keiser et al., 2016).

GMAT

The degree to which the relationship between GMAT scores and graduate school performance varies for relevant student subgroups has been the focus of numerous research studies. For example, in their 2007 meta-analysis, Kuncel et al. tested whether the relationship between GMAT scores and first-year graduate GPA was moderated by gender, native language, full-time status, and undergraduate major. The results indicated that the predictive strength of GMAT scores did not vary by student subgroups except for a slightly stronger relationship for part-time students.

More recently, Talento-Miller (2017) examined the differential validity and differential prediction of GMAT scores in terms of graduate GPA by gender, race/ethnicity, citizenship, native language, and age. Based on data from 28 business programs, the results indicated that the predictive strength of GMAT Total scores is similar across subgroups, with slightly higher correlations for older students (≥31) and slightly lower

correlations for non–native English speakers and students who were citizens of East or Southeast Asia. In general, the results support the predictive validity of GMAT scores, overall and for student subgroups.

GRE

Prior research has also examined GRE's prediction by common subgroups such as gender and ethnicity. For example, Burton and Wang (2014) examined GRE's prediction of first-year GPA and faculty ratings in the mastery of the discipline, professional productivity, and communication skills across five disciplinary areas (biology, chemistry, education, psychology, and English). Results did not show consistent over- or underprediction for gender groups. Also, after course taking was controlled within a discipline, the gender prediction differences largely disappeared. The findings showed sizable overprediction for African Americans. However, the authors speculated that the sample size for this group may have had an impact on such conclusions. Kuncel and Hezlett (2007) reported that underprediction for females was observed at the undergraduate level but not at the graduate level and cited differential disciplinary choice and course selection as possible reasons accounting for the differential gender prediction.

LSAT

In addition to examining the overall validity of LSAT scores for predicting first-year GPA among law students, the LSAC also regularly investigates the degree to which LSAT scores exhibit differential prediction of first-year law school GPA by relevant subgroups, alone and in combination with UGPA. Two reports were recently published investigating this issue by gender and racial/ethnic subgroups using data from schools participating in the LSAT Correlation Studies from 2012 to 2015 (Rahman & Trierweiler, 2017; Trierweiler & Rahman, 2017). Encompassing data from 148 law schools with sufficient racial/ethnic sample sizes, the results indicated the LSAT first-year GPA validity coefficients were similar across racial/ethnic subgroups (Trierweiler & Rahman, 2017). Both LSAT scores and UGPA overpredicted first-year GPA for underrepresented racial/ethnic minority students, with the combination of the two measures resulting in the least amount of differential prediction.

As for the analyses by gender (Rahman & Trierweiler, 2017), the results were based on data from 158 law schools. Similar to the racial/ethnic findings, the results indicated the LSAT first-year GPA validity coefficients were similar for males and females. In terms of differential prediction, LSAT scores, alone and in combination with UGPA, exhibited minimal prediction error by gender, supporting the validity and fairness of LSAT scores for use in law school admission decisions.

MCAT

Prior studies have examined the degree to which MCAT scores exhibit differential validity, with results indicating comparable validity across racial/ethnic subgroups (Busche

et al., 2020). However, with the introduction of the new MCAT in 2015, the collection of evidence supporting the validity and fairness of new MCAT scores is essential. In response to that research gap, Busche and colleagues (2020) not only investigated the overall predictive validity of the new MCAT scores but also examined whether new MCAT scores exhibited differential prediction of first-year medical GPA and on-time progression to Year 2 by demographic characteristics (gender, race/ethnicity, and highest parental education level). Based on an analysis of mean residuals, there was no evidence of differential prediction by the subgroups examined; none of the mean residuals were significantly different from zero.

PLACEMENT TESTS

Placement Decisions: Remediation Versus Credit-Bearing Courses

According to the *Standards* (AERA et al., 2014), placement tests are designed to be efficient and usually cover only a subset of prerequisite content taught in high school. Postsecondary institutions may use scores to determine student eligibility to enter credit-bearing courses rather than remedial developmental courses or to determine the appropriate initial course in a sequenced area of study (e.g., English, foreign language, mathematics).

Historically, placement in remedial or college-level courses has been based on a single placement test score, often locally determined through a study or faculty judgment. Remedial courses may be helpful to some students, but they also require students to make a substantial investment of time and money that could otherwise be applied to college-level coursework. Studies have suggested that the effects of remedial courses on student outcomes are at best mixed for those who are thought to be on the cusp of needing additional academic support. Students who enroll in one or more remedial courses must still pay tuition without earning credits toward graduation and remedial coursework may exhaust students' financial aid before they graduate, reducing their graduation rates (Attewell et al., 2006). Colleges also incur a large financial burden from developmental programs, estimated at more than $4 billion in 2015 (Rodriguez et al., 2015).

The high cost in dollars and opportunity associated with remedial courses spurred legislation and policies that sought to limit remediation by eliminating remedial courses at some institutions (e.g., 4-year colleges), require evidence of readiness for college admissions (e.g., ACT/SAT or placement test scores), or require the use of multiple measures for placement decisions. While the number and percentage of students placed into remediation show a moderate decrease since 2010, nearly 70% of new students entering a 2-year college each fall are assigned to such programs (Chen & Simone, 2016; Rodriguez et al., 2015).

Bailey et al. (2010) examined the enrollment patterns of 257,672 first-time credential-seeking students who began their enrollment in fall 2003 to fall 2004 in 57 colleges that participated in a developmental program entitled "Achieving the

Dream." Student enrollments and course completions were tracked for 3 years. The researchers found that about 30% of students referred to developmental education did not enroll in any remedial course, and only about 60% of referred students enrolled in the remedial course to which they were referred. The results also show that more students exit their developmental sequences because they did not enroll in the first or a subsequent course than because they failed or withdrew from a course in which they were enrolled. Bailey et al. (2010) also found that students who enrolled in a college-level class despite a recommended placement in remedial courses were slightly less likely to pass the course than students who were placed directly into the college-level course by virtue of passing the placement test but were much more successful in the course than students who complied with their remedial placement, because relatively few students who entered remediation ever attempted the college-level course. Despite the prevalence of remedial programs in colleges, studies of program effectiveness and the validity of placement tests assigning students to such programs have produced at best mixed results, with most studies suffering from small samples and a lack of rigor (Scott-Clayton, 2012). A more comprehensive study of remediation explored the interaction of remediation and academic preparation, finding that students who exhibited poor precollege academic preparation (as measured by high school grades, the highest math course taken in high school, and college admission test scores) and completed remedial programs fared better on a range of college outcomes than similarly prepared students who did not complete remediation programs; however, these findings did not hold for students who had moderate or strong precollege academic preparation and were assigned to remedial courses (Chen & Simone, 2016).

In 2012, the National Assessment Governing Board commissioned a college survey to determine which placement tests and cut scores were most prevalent. The study documents differences across 2-year and 4-year institutions (Fields & Parsad, 2012). Table 17.7 illustrates that 80% of 2-year institutions use a placement test in math compared to 65% of 4-year institutions. ACT math was used at about a quarter of institutions (2-year and 4-year), whereas the SAT was being used at 20% of 4-year institutions but only 12% of 2-year institutions. Seventy-three percent of 2-year institutions reported using a placement test in reading as compared to less than 40% of 4-year institutions. ACT's Compass and Asset and College Board's ACCUPLACER placement tests were the most popular instruments at 2-year colleges, but both ACT tests were discontinued because ACT cited research showing placement tests, in general, are not as effective at determining readiness as previously thought (Fein, 2015). Of course, these data are now over 10 years old, and patterns in usage may have changed.

Given that test scores tend to be only moderately predictive of course success, concerns about the degree to which test scores lead to incorrect placement decisions have surfaced, especially in light of the poor educational outcomes for students placed in remedial college courses (Barnett & Reddy, 2017). Placement tests have been found

Table 17.7 Estimated Percentage of U.S. Postsecondary Institutions Using Selected Math and Reading Tests for Placement in Fall 2011

Institutional Level	Est. Number of Institutions	% Using Any Math Test	ACT Math	SAT Math	ACCUPLACER Elementary Algebra	ACCUPLACER College Math	Compass Algebra	Compass College Algebra	Other Math
All institutions	3,650	71	23	17	16	5	20	4	22
2-year	1,470	80	23	12	24	7	34	8	16
4-year	2,180	65	24	20	11	3	10	2	26
Institutional level	Estimated Number of Institutions	% Using Any Reading Test	ACT Reading	SAT Critical Reading	ACCUPLACER Reading Comprehension	ASSET Reading Skills	Compass Reading	Other Reading	
All institutions	3,650	53	16	11	19	9	22	10	
2-year	1,470	73	21	10	29	19	43	12	
4-year	2,180	39	13	12	12	2	9	8	

Note. The number of institutions may not sum to totals because of rounding or suppressed data. Adapted from *Tests and Cut Scores Used for Student Placement in Postsecondary Education: Fall 2011* (p. 11, Table 1, and p. 17, Table 6), by R. Fields and B. Parsad, 2012, National Assessment Governing Board, U.S. Department of Education.

to underplace students, in comparison to coursework completed and grades, having an adverse impact on underrepresented minorities. In response, an emerging trend in college placement is for institutions to take a more holistic approach (Rutschow & Mayer, 2018), using additional measures such as high school grades, high school courses completed, and performance in dual enrollment or Advanced Placement grades. Moreover, legislative and policy eliminating the use of placement tests have been associated with large increases in retention, transfer credits, and graduation rates at 2-year colleges (Zamundio-Suarez, 2022). About 13% and 21% of all institutions report using one of these measures along with test scores for placement decisions in reading and math, respectively (Fields & Parsad, 2012).

ACCUPLACER

College Board's ACCUPLACER placement testing program is the only remaining major large-scale placement test after ACT discontinued both Compass and Asset placement tests in 2016. For over 30 years, ACCUPLACER (originally branded as Computer Placement Tests) has been used to assess student readiness for college-level courses. The ACCUPLACER battery of placement tests, labeled Next Generation ACCUPLACER, was redesigned in 2016 based on data from a survey of higher education users. New items were written based on new specifications and field tested and calibrated to a new reporting scale, 200–300, with increments of 1. The battery includes tests of English Literacy (EL) and tests of computer skills, but this chapter will focus only on the Next-Generation placement tests: Reading (20 questions), Writing (25 questions), Arithmetic (20 questions), Quantitative Reasoning, Algebra, and Statistics (20 questions), and Advanced Algebra and Functions (20 questions).

ACCUPLACER has also been used with high school students to identify students prepared to enter college credit–bearing courses or in need of further developmental instruction prior to graduation. The technical manual states that scores can be used to identify high school students "eligible" to enroll in dual enrollment courses but provides no evidence or guidance on such uses (College Board, 2019b). All ACCUPLACER assessments are computer adaptive, fixed length, and tailored to each test taker using an item-selection algorithm based on a weighted deviations model.[13] A number of constraints are built into the assessments, including content constraints. In 2016, 11 million ACCUPLACER tests were administered, but that number counts each assessment completed; the same student will often be routed through multiple tests in a single subject (e.g., all three math or EL tests) to receive a single placement score (College Board, 2017c, 2019b); additionally, volume counts are based on the larger set of assessments used, including Next-Generation, traditional, EL, computer skills, write placer, and diagnostic tests.

The program handbook states, "Multiple factors should be used when making high-stakes decisions like college placement" (College Board, 2018, p. 4) and encourages colleges to conduct a local validity study if establishing cut scores. The Next-Generation

ACCUPLACER redesign incorporated changes in content, item types, and other elements, but publicly available statistical specifications are based on the prior design (College Board, 2017c).

Technical Characteristics

Reliability estimates range from .80 to .92, and standard errors of measurement range from about 7 to 12 points on the 200–300 point scale (L. Bay, personal communication, April 27, 2020). The ACCUPLACER Battery relies on content validity to support claims and provides a detailed description of content and skills measured. Although criterion-related (e.g., predictive) studies have not been reported for ACCUPLACER Next-Generation, the previous battery had moderately strong evidence supported by a meta-analysis of 47 placement validity studies across 17 institutions between 2001 and 2006 (Mattern & Packman, 2009). The mean operational validity coefficient for individual ACCUPLACER tests in the previous battery (before 2016) ranged from .16 to .36 (M = .24) when success was defined as obtaining a "B or higher" and from .10 to .32 (M = .19) when success was defined as obtaining a "C or higher," suggesting a weak to moderate relationship between ACCUPLACER scores and course success. However, mean validity coefficients increased to .42 and .34 after correcting for restriction of range. Validity of placement tests is also evidenced by examining the percentage of students recommended for placement in a course who eventually pass that course. Such studies have shown moderate to strong relationships for the previous tests.

STUDENT LEARNING OUTCOMES

Definition of SLO Assessment

SLO have received heightened attention from many higher education institutions for reasons such as accreditation, accountability, and internal improvement (Liu, 2023). The National Institute for Learning Outcomes Assessment (NILOA; 2019), a pioneer in the promotion of valid use and effective practices of SLO assessment, defines learning outcomes as "the expected knowledge, skills, attitudes, competencies, and habits of mind that students are expected to acquire from an educational experience (para. 1)." It is important to note that learning outcomes encompass both domain-specific knowledge and skills and generic, transferable competencies.

In the acquisition of domain-specific knowledge and skills (e.g., journalism or economics), students typically follow a structured progression of a curriculum, designed purposefully to build and reinforce their knowledge and skills in these domains. In this regard, students' increased understanding and proficiency can be largely attributed to their experience in that defined sequence of courses.

Students' development of generic, transferable competencies such as critical thinking and intercultural competency is more complicated and can hardly be attributed to a single college experience. For example, institutions may not offer critical thinking

courses but still include critical thinking as a measured outcome, since students could gain critical thinking skills from exposure to a broad range of courses, programs, and institutional activities. Students may hone critical thinking skills, for example, from a philosophy course that discusses inductive and deductive reasoning and from a statistics course that introduces correlation and causation. Students could also gain critical thinking skills from participating in an undergraduate research program or being part of a debate club. Furthermore, they could take a Massive Open Online Course (MOOC) on critical thinking. All these activities can contribute to a student's critical thinking skills without the student's enrollment in a course dedicated to critical thinking. In this sense, students' critical thinking skills reflect the overall learning experience they have at an institution.

In addition to the important distinction between domain-specific and generic learning outcomes, it is also critical to note that the focus of college learning outcomes by higher education institutions is on a spectrum of cognitive abilities, personality and behavioral traits, and social and emotional skills (Pellegrino & Hilton, 2012). The focus is closely aligned with the "whole student" notion and also reflects the demands for both technical and nontechnical skills in the modern workplace. A notable example is the three clusters of competencies identified: cognitive (e.g., critical thinking, creativity), intrapersonal (e.g., intellectual openness, work ethics), and interpersonal (e.g., collaboration, leadership). Moreover, the value placed on nontechnical skills by employers has increased over time. The 2018 IBM Institute for Business Value Global Country Survey revealed that adaptability, time management and prioritization, teamwork, and communication overtook technical skills in becoming the most valued skills by employers (LaPrade et al., 2019).

Notable International SLO Efforts

In parallel to the attention SLO assessment has received in the United States, there are also a number of notable international initiatives about the promotion of college-level learning outcomes. For example, the Assessment of Higher Education Learning Outcomes (Tremblay et al., 2012) project, sponsored by the Organisation for Economic Co-operation and Development (2013), attracted 17 countries to participate in its pilot examining college students' performance in critical thinking, quantitative reasoning, and economics, reflecting a focus on both generic and domain-specific skills.

Another notable example is the Modeling and Measuring Competencies in Higher Education–Validation and Methodological Innovations (KoKoHs) project, funded by the German Federal Ministry of Education and Research over two phases, totaling 9 years (2011–2015; 2015–2019). KoKoHs was focused on the assessment of competencies in both general and disciplinary-specific domains and built a consortium of institutions and programs in Germany and with its international partners in addressing students' learning outcomes (Zlatkin-Troitschanskais et al., 2018).

The European Union also funded a two-phase project titled Measuring and Comparing Achievements of Learning Outcomes in Higher Education in Europe to identify

a common assessment framework for five broad areas of disciplinary domains: engineering/civil engineering, social sciences/education, humanities/history, health care/nursing, and natural sciences/physics. On the basis of collaborations between faculty members and with input from ETS, a series of assessment frameworks were published articulating the knowledge, skills, and competencies students are expected to master and demonstrate at both the bachelor and the master levels (Wagenaar, 2018).

Different Types of SLO Assessments

The majority of U.S. higher education institutions specify their own learning objectives and student learning outcomes. According to a survey of 811 regionally accredited, undergraduate degree-granting institutions, Jankowski et al. (2018) reported that 82% of institutions have specified learning outcomes. The institutions surveyed reported using multiple assessment approaches to measure learning outcomes using four different types of tools on average. The most popular assessment tools include national student surveys, alumni surveys, locally developed surveys, standardized assessments of general knowledge and skills, rubrics that define competencies (e.g., VALUE Rubrics by Association of American Colleges & Universities), employer surveys, external performance assessments, and portfolios. Notable examples of standardized SLO measures include the HEIghten suite of assessments (Liu et al., 2014), the Collegiate Learning Assessment (Kaniuka & Wynne, 2019), and the California Critical Thinking Assessment (Davies, 2011). The HEIghten assessment suite includes five independent modules in critical thinking, written communication, quantitative literacy, civic competency, and intercultural competency (Griffith et al., 2016; Liu et al., 2014; Roohr et al., 2014; Sparks et al., 2014; Torney-Purta et al., 2015). Rubrics are also a popular tool used by faculty members to evaluate learning outcomes. Each assessment tool has its own advantages and limitations. For example, national student surveys are easy to implement and can be an efficient way of gathering information on student experience and engagement on campus, yet they typically do not provide direct evidence of student learning. Rubrics offer great flexibility for adaptation but could fall short of consistency in implementation if users are not adequately trained. As Liu (2017) pointed out, there is no one-size-fits-all solution when it comes to assessing student learning outcomes. Depending on their assessment goals, priorities, and resources, institutions should select the tools that best fit their needs.

How SLO Assessments Have Been Used and Current State

SLO assessments serve many purposes, such as a program or regional accreditation, internal institutional improvement, faculty/staff interest, equity concerns, and responding to accountability calls (Jankowski et al., 2018). The assessment results are typically reported at an aggregate level. As noted earlier, the purpose of the Commission on the Future of Higher Education was to identify prominent issues in American higher education and develop improvement solutions. The commission's first report (U.S. Department of Education, 2006) sparked a series of accountability initiatives that required participating institutions to provide more data on key performance metrics

to the public, with particular attention to the assessment of learning outcomes using standardized measures. The Department of Education also provided a grant with a component to examine the comparability of students' performance in critical thinking and analytic writing assessed across three standardized measures (Klein et al., 2009). National accountability efforts include the Voluntary System of Accountability for public 4-year institutions and the Voluntary Framework of Accountability created for community colleges.

As is true for many initiatives and changes in education, the focus on accountability has evolved since 2012 as a result of challenges in implementation and shifting policy guidance. The impact of accountability initiatives on institutions' use of learning outcomes assessment has declined significantly, from 31% in 2007 to 13% in 2017 (Jankowski et al., 2018). Internal improvement has become an increasingly important driver for institutions to examine students' learning. Some of the most significant challenges that institutions face include decreasing student enrollment, rising student debt, competition from nontraditional education providers (e.g., online programs, boot camps), and lack of connection to the workforce. When asked whether most Americans have an accurate view of higher education, only 15% of college and university presidents agreed or strongly agreed (Inside Higher Ed, 2019). For many institutions, the focus on SLOs has been driven by the need to prove value instead of a response to federal calls for accountability.

Prominent Issues in Assessing Learning Outcomes

Test-Taking Motivation

In this section, we discuss a number of issues in the implementation and interpretation of SLO assessments. As previously mentioned, most institutions administer SLO assessment to drive institutional change, rather than to monitor individual learning. The assessment outcomes rarely bear direct consequences for students. As a result, students' motivation for taking SLO assessments has been a long-standing concern. Abundant research suggests that when test stakes are low, test takers may not try their best; and therefore, their performance tends to be an inaccurate reflection of their abilities (e.g., Braun et al., 2011; Gneezy et al., 2019; Liu et al., 2012; Wise & DeMars, 2005). Gneezy et al.'s (2019) study revealed that if U.S. students had used their best effort in the Programme for International Student Assessment (PISA), U.S. ranking on the PISA mathematics test would have been improved from 36th to 19th. Researchers have explored strategies to help improve students' motivation. Offering incentives contingent on performance seems to be an effective way of improving motivation and performance (e.g., Braun et al., 2011; Gneezy et al., 2019).

Measuring Performance Differences

A concept that was frequently explored in the beginning of the accountability initiatives around 2007 was value-added learning of college education, implying that the amount of learning that can be attributed to the effectiveness of higher education could be

accurately measured. Although the concept was appealing in holding education providers accountable for the instruction and experiences they provide, there were many designs as well as methodological and practical constraints that prevented meaningful interpretations. In the analysis of change of learning, longitudinal samples are typically preferred over cross-sectional samples because of the invariance in key background variables for longitudinal samples. However, longitudinal samples are very difficult to obtain in higher education because the first-year to second-year attrition rate is about 17% for 4-year public institutions and 38% for 2-year colleges (National Student Clearinghouse Research Center, 2019). In addition, significant transfer rates also make it difficult to track students. About 37% of students who start at 2-year institutions and about 39% of students who start at 4-year institutions transfer to another institution within 6 years (Shapiro et al., 2018). These are the main reasons that many studies employ cross-sectional samples when examining learning gains in college (Klein et al., 2009; Liu et al., 2012; Roohr et al., 2019). However, students in different cohorts may vary in key background and academic variables, which could potentially influence the comparability of these groups (Liu, Liu, et al., 2016). For example, senior students who manage to remain in college may have possessed higher academic qualifications on average when they were freshmen than the current freshmen in the cross-sectional sample, thus introducing inequality in the two samples to be compared. There is another issue of differential participation from freshmen and seniors in that seniors are less likely to sign up for voluntary institutional testing because of other priorities, which further exacerbates the sample inequality. The confluence of methodological challenges, differential motivation (e.g., freshmen try harder than seniors), and mobility in higher education poses difficulty on any attempt to disentangle the contribution of any individual institution on student learning. Researchers have also provided solutions to address unequal samples when examining cross-sectional learning gain. For example, Liu, Liu, et al. (2016) explored the application of a propensity scoring weighting approach aiming to control for possible differences in key background variables (e.g., gender, ethnicity, college entrance exam scores) between cohorts of freshmen and seniors used in the analysis.

Reimagining Learning Outcomes Assessment: Connecting Assessment, Learning, and Careers

Despite the widespread use of learning outcomes assessments, institutions have varying degrees of success in utilizing the assessment results for driving improvement and innovation. A unique challenge around the interpretation and use of generic SLO assessments is that the results are typically not directly related to any course but instead reflect a culmination of a wide range of experiences in and outside college. Institutions at times struggle with identifying a practical entry point to provide remedies when students show a deficiency in a generic field. According to Jankowski et al. (2018), there are many examples of expansive assessment efforts that lead to limited changes on campus. In the following section, we discuss how the use of SLO assessments can be reimagined to help enhance students' key competencies and increase connection with the workforce.

Employers have long been concerned about skills gaps, and such concerns are exacerbated by industry changes introduced by automation. McKinsey Global Institute (2017) projected that by 2030, about 3%–14% of the global workforce will need to switch occupational categories, and all workers will need to adapt as their jobs evolve with automation. An employer survey from the National Association of Colleges and Employers (2018) suggested significant gaps in key competency areas, including critical thinking/problem-solving, global/multicultural fluency, career management, digital technology, leadership, teamwork and collaboration, oral and written communication, and professionalism and work ethic. These skills are also emphasized in Liu, Kell, Liu, et al. (2023) in which the researchers proposed a new vision for assessment that reflects the continuity of skills development throughout K–12, higher education, and workforce. Although institutional assessment reflects a wide range of educational objectives, having a connection to employer expectations has great potential for increasing students' workforce readiness and engaging students in the assessment process because the results will likely bear more significance for students. Offering a clear definition of what these skills mean and how they manifest in concrete contexts is also important as institutions strive to help students improve on these critical skills, which calls for close collaboration between education providers and employers to contextualize these skills (Liu et al., 2021; Liu, Kell, Williams, et al., 2023). Currently, learning resources for skill improvement are provided from a range of sources, including locally or institutionally developed resources MOOCs, and offerings from education start-ups and non-traditional education programs. Making students aware of these resources and offering more systematic support following the assessment activities has great potential to motivate students to capitalize on the learning opportunities for skill improvement.

CONCLUSION

The use of tests in higher education, especially for admissions, remains one of the most highly contested uses of standardized testing in America. College degrees remain a highly valued commodity in our society. Entrance to selective institutions is valued more than entrance to typical or less selective institutions, and the ratio of applicants to openings is highly restricted in the most selective undergraduate and many graduate and professional degree programs. As long as these conditions exist, there will be heated debate about who gets in and the tools and criteria used to make such decisions. As Zwick (2017) reminded us, there is no universally accepted definition of merit or equity and no completely objective way to evaluate admissions systems or decisions. The outcomes of admissions processes serve as prima facie evidence of a lack of fairness for many stakeholders, and accumulated evidence of validity, reliability, and fairness of scores is often unpersuasive.

However, results of the body of research on admissions tests continue to reflect their value and efficacy, as described by a 1999 report from the National Academy of Sciences. This study (Beatty et al., 1999) noted that standardized tests provide an efficient, low-cost, valid, and reliable method to identify talent (college readiness) and minimize

differences that exist between schools, grades, and learning opportunities. A more recent study found that admissions tests result in greater diversity, and policies that would eliminate or reduce their use would reduce diversity and have other negative effects on educational outcomes (University of California, Academic Senate, 2020). Although this finding was not directly rejected, the UC president and Regents did reject the recommendations from the task force. However, admission tests must also be used appropriately, which precludes reliance on test scores as the primary or sole measure used for admissions, placement, or assessment of learning outcomes. The discrepancy between the public's perception of standardized tests and the research that contradicts common beliefs points to the need for the measurement community to do a more effective job in promoting a national dialogue of assessment that is data driven and evidence based.

Another important trend emerging in the use of assessments in higher education includes the increasing influence of technology because all major admissions test are now digital and models of testing such as remote proctoring and increased use of artificial intelligence have emerged. In turn, the marked demands for increased flexibility and customization may challenge the very principles associated with standardized testing and score equivalence (Camara & Davis, 2022). Similar to K–12 state assessments in accountability, four of the major admissions tests discussed have also undertaken substantial design changes to reduce testing time and length. Reducing testing time, stress of anxiety of test takers, and providing free and substantial programs for test preparation have gained importance in this industry segment. We should anticipate continued efforts to improve the test taker experience, given its heightened importance in a test optional environment where learners have much greater autonomy and flexibility in their testing choices.

In the brief period between completion of the first draft of this chapter (late 2019) and the final version (2025), testing in higher education has undergone the most profound changes in its history. These changes, continued challenges, and future uses are still very fluid and unknown as we conclude. In the next few years, we anticipate the release of additional research, which can better help us evaluate the efficacy and utility of admissions testing, and we hope such research will also address the efficacy and consequences of admissions practices and decisions made based on other factors. Placement testing has undergone additional scrutiny and the assessment of learning outcomes is in its infancy. The next decade will present challenges to testing organizations and higher education in both continuing practices of assessing students for admissions and placement and deriving meaning from assessments of student learning.

ACKNOWLEDGMENTS

We would like to thank the editors, Linda Cook and Mary Pitoniak, for their guidance and coordination of this work and especially our reviewers, Jerry Lucido, Emily Shaw, and Rebecca Zwick, for their detailed review, suggestions, and comments over multiple reviews, which only strengthened the manuscript.

REFERENCES

ACT. (2009). *ACT: The first fifty years, 1959–2009*.

ACT. (2018). *ACT/SAT concordance*. https://www.act.org/content/dam/act/unsecured/documents/pdfs/ACT-SAT-Concordance.pdf

ACT. (2019a). *The ACT Profile Report—National graduating class 2018*. https://www.act.org/content/dam/act/unsecured/documents/cccr2018/P_99_999999_N_S_N00_ACT-GCPR_National.pdf

ACT. (2019b, October 8). *ACT test to provide new options designed to offer students more choices, faster results*. http://leadershipblog.act.org/2019/10/act-test-to-provide-new-options.html

ACT. (2024a, January). *ACT Assessment technical manual*. https://www.act.org/content/dam/act/unsecured/documents/ACT_Technical_Manual.pdf

ACT. (2024b). *ACT Test enhancements and changes*. https://www.act.org/content/act/en/products-and-services/the-act/test-changes/enhancements.html

Adams, C. J. (2017, May 30). In race for test-takers, ACT outscores SAT—for now. *Education Week*. https://www.edweek.org/ew/articles/2017/05/24/in-race-for-test-takers-act-outscores-sat-for.html

Allen, J. (2006). *The benefits of statewide use of the ACT Test*. ACT. https://files.eric.ed.gov/fulltext/ED510717.pdf

Allen, J. (2013). *Updating the ACT college readiness benchmarks* (Report No. 2013-6). ACT. https://www.act.org/content/dam/act/unsecured/documents/ACT_RR2013-6.pdf

Allen, J. (2016). *Subgroup representation and validity in ACT College Readiness Benchmark development samples*. ACT. https://www.act.org/content/dam/act/unsecured/documents/5483-tech-brief-subgroup-representation.pdf

Allen, J., & Cruce, T. (2025, June). *Initial evidence supporting interpretations of scores from the enhanced ACT test* (ACT Research Report). ACT. https://www.act.org/content/dam/act/unsecured/documents/R2425-Initial-Evidence-Supporting-Interpretations-of-Scores-from-the-Enhanced-ACT-Test-2025-06.pdf

Allen, J., & Sconing, J. (2005). *Using ACT assessment scores to set benchmarks for college readiness*. (Report No. RR2003-1). ACT. https://www.act.org/content/dam/act/unsecured/documents/ACT_RR2005-3.pdf

Alvero, A. J., Glebel, S., Gebre-Medhin, B., Antonio, A. L., Stevens, M. L., & Dominque, B. W. (2021). Essay content and style are strongly related to household income and SAT scores: Evidence from 60,000 undergraduate applications. *Science Advances*, *7*(42), 1–10. https://doi.org/10.1126/sciadv.abi9031

American Bar Association. (2021, November 19). *Summary of actions of the section's council at its public meeting Nov. 19, 2021*. https://www.americanbar.org/content/dam/aba/administrative/legal_education_and_admissions_to_the_bar/council_reports_and_resolutions/nov21/21-nov-legal-ed-summary-of-council-meeting.pdf

American Educational Research Association, American Psychological Association, & National Council on Measurement in Education. (2014). Reliability/precision and errors of measurement. In *Standards for educational and psychological testing* (pp. 33–47). https://www.testingstandards.net/uploads/7/6/6/4/76643089/9780935302356.pdf

Anderson, N. (2014, August 20). Growing number of states fund ACT college admissions testing for 11th-grade students. *Washington Post*. https://www.washingtonpost.com/local/education/growing-number-of-states-fund-act-college-admission-testing-for-11th-grade-students/2014/08/19/eb0edcba-27a7-11e4-86ca-6f03cbd15c1a_story.html

Anthony, L. C., Dalessandro, S. P., & Reese, L. M. (2013). *Predictive validity of the LSAT: A national summary of the 2011 and 2012 LSAT correlation studies* (LSAT Technical Report No. TR-13-03). Law School Admission Council.

Anthony, L. C., Dalessandro, S. P., & Trierweiler, T. J. (2016). *Predictive validity of the LSAT: A national summary of the 2013 and 2014 LSAT Correlation Studies* (LSAT Technical Report No. TR-16-01). Law School Admission Council.

Arundel, K. (2024, April 18). *ACT partners with private equity firm, will transition to for-profit*. Higher Ed Dive. https://www.k12dive.com/news/act-assessment-private-equity-college-admissions/7134450/

Association of American Medical Colleges. (2015). *The MCAT exam: Year at a glance 2015*.

Association of American Medical Colleges. (2022). *AAMC PREview professional readiness exam*. https://students-residents.aamc.org/aamc-preview/aamc-preview-professional-readiness-exam

Association of American Medical Colleges. (2023a). *Table A-10: Applicants to U.S. MD-Granting Medical Schools by Race/Ethnicity (Alone) and State of Legal Residence*, 2023–2024. https://www.aamc.org/media/6036/download?attachment

Association of American Medical Colleges. (2023b). Table A-11: Matriculants to U.S. MD-Granting Medical Schools by Race/Ethnicity (Alone) and State of Legal Residence, 2023–2024. https://www.aamc.org/media/6041/download?attachment

Association of American Medical Colleges. (2025a). *MCAT exam scoring*. https://students-residents.aamc.org/register-mcat-exam/publication-chapters/mcat-exam-scoring

Association of American Medical Colleges. (2025b). *What's on the MCAT exam? Exam overview*. https://students-residents.aamc.org/whats-mcat-exam/publication-chapters/whats-mcat-exam

Association of American Medical Colleges. (2025c). *What's on the MCAT exam? PDF Outline*. https://students-residents.aamc.org/prepare-mcat-exam/whats-mcat-exam-pdf-outline

Attewell, P. A., Lavin, D. E., Domina, T., & Levey, T. (2006). New evidence on college remediation. *Journal of Higher Education*, *77*(5), 886–924. https://doi.org/10.1080/00221546.2006.11778948

Bailey, T., Jeong, D. W., & Cho, S. W. (2010). Referral, enrollment, and completion in developmental education sequences in community colleges. *Economics of Education Review*, *29*(2), 255–270. https://doi.org/10.1016/j.econedurev.2009.09.002

Barnett, E., & Reddy, V. (2017). College placement strategies: Evolving considerations and practices. In K. L. McClarty, K. D. Mattern, & M. N. Gaertner (Eds.), *Preparing*

students for college and careers: Theory, measurement, and educational practice (pp. 82–93). Routledge.

Bauer-Wolf, J. (2023, February 7). NCAA permanently ends SAT, ACT eligibility requirement for Division I, II student-athletes. *Higher Ed Dive*. https://www.highereddive.com/news/ncaa-permanently-ends-sat-act-eligibility-requirement-division-i-ii/642117/

Beatty, A., Greenwood, M. R. C., & Linn, R. L. (Eds.). (1999). *Myths and tradeoffs: The role of tests in undergraduate admissions*. National Academies Press.

Belasco, A. S., Rosinger, K. O., & Hearn. J. C. (2015). The test-optional movement at America's selective liberal arts colleges: A boon for equity or something else? *Educational Evaluation and Policy Analysis*, *37*(2), 206–223. https://doi.org/10.3102/0162373714537350

Bennett, C. T. (2022). Untested admissions: Examining changes in application behaviors and student demographics under test-optional policies. *American Educational Research Journal*, *59*(1), 180–216. https://doi.org/10.3102/00028312211003526

Berry, C. M., & Sackett, P. R. (2009). Individual differences in course choice result in underestimation of the validity of college admissions systems. *Psychological Science*, *20*(7), 822–830. https://doi.org/10.1111/j.1467-9280.2009.02368.x

Bingham, R. O., Chanatry, J., Baffi-Dugan, C., Childress, B., & Maxwell, S. A. (2012). Preparing for the new MCAT: The perspective of advisors. *Peer Review*, *14*(4), 25–29.

Black Lives Matter: About. (2024). *Black lives matter*. http://blacklivesmatter.com/about/.

Boost, I. (2019, March 21). What the scammers got right about college admissions. *The Atlantic*. https://www.theatlantic.com/education/archive/2019/03/real-college-bribery-scandal-whats-legal/585298/

Boshara, R., Emmons, W. R., & Noeth, B. J. (2015). The demographics of wealth—how age, education and race separate thrivers from strugglers in today's economy. *Essay No. 2: The Role of Education.* . Center for Household Financial Stability, Federal Reserve Bank of St. Louis. https://www.stlouisfed.org/~/media/Files/PDFs/HFS/essays/HFS-Essay-2-2015-Education-and-Wealth.pdf

Bowen, W. G., & Bok, D. (1998). *The shape of the river: Long-term consequences of considering race in college and university admissions*. Princeton University Press.

Braun, H., Kirsch, I., & Yamamoto, K. (2011). An experimental study of the effects of monetary incentives on performance on the 12th grade NAEP reading assessment. *Teachers College Record*, *113*, 2309–2344. http://dx.doi.org/10.1177/016146811111301101

Bridgeman, B. (2016). Can a two-question test be reliable and valid for predicting academic outcomes? *Educational Measurement: Issues and Practice*, *35*(4), 21–24. https://doi.org/10.1111/emip.12130

Bridgeman, B., McCamley-Jenkins, L., & Ervin, N. (2000). *Predictions of freshman grade-point average from the revised and recentered SAT® I: Reasoning Test* (College Board Research Report No. 00-1). ETS. https://doi.org/10.1002/j.2333-8504.2000.tb01824.x

Burton, N. W., & Wang, M. M. (2014). *Predicting long-term success in graduate school: A collaborative validity study*. In C. Wendler, B. Bridgeman, & C. Ezzo, C. (Eds.), *The research foundation for the GRE revised General Test: A compendium of studies* (pp. 5.6.1–5.65). ETS. https://www.ets.org/s/research/pdf/gre_compendium.pdf

Busche, K., Elks, M. L., Hanson, J. T., Jackson-Williams, L., Manuel, R. S., Parsons, W. L., Wofsy, D., & Yuan, K. (2020). The validity of scores from the new MCAT exam in predicting student performance: Results from a multisite study. *Academic Medicine, 95*(3), 387–395. https://doi.org/10.1097/acm.0000000000002942

Camara, W. J. (2009). College admission testing: Myths and realities in an age of admissions hype. In R. Phelps (Ed.), *Correcting fallacies about educational and psychological testing* (pp. 147–180). American Psychological Association.

Camara, W. J. (2018). *Illustrating differences among students with the same HSGPA of 4.0 or above* [Unpublished manuscript]. ACT.

Camara, W. J. (2020). Never let a crisis go to waste: Large scale assessment and the response to COVID-19. *Educational Measurement: Issues and Practice, 339*(3), 10–18. https://doi.org/10.1111/emip.12358

Camara, W. J. (2023). Admissions testing impact on access and alternative opinions. In K. Geisinger (Ed.), *College admissions testing* (pp. 10–31). Buros Institute.

Camara, W. J. (2024). Admission testing in higher education: Changing landscape and outcomes from test-optional policies. *Educational Measurement: Issues and Practice, 43*, 104–111.

Camara, W. J., & Davis, L. (2022). Fairness concerns resulting from innovations and applications of technology to assessment. In J. Jonson & K. Geisinger (Eds.), *Fairness issues and solutions in educational and psychological testing: Implications for researchers, practitioners, policy makers, and the public* (pp. 153–187). American Educational Research Association.

Camara, W. J., Kimmel, E., Scheuneman, J., & Sawtell, E. (2003). Whose grades are inflated? (*CollegeBoard Research Report 2003–4)*. CollegeBoard. http://professionals.collegeboard.com/profdownload/pdf/04843cbreport20034_31757.pdf

Camara, W. J., & Mattern, K. (2022). Inflection point: The role of testing in admissions decisions in a post pandemic environment. *Educational Measurement: Issues and Practice, 41*, 10–15. https://onlinelibrary.wiley.com/doi/epdf/10.1111/emip.12493

Camara, W. J., Mattern, K., Croft, M., Vispoel, S., & Nichols, P. (2019). A validity argument in support of the use of college admission test scores for federal accountability. *Educational Measurement: Issues and Practice, 38*(4), 12–26. https://doi.org/10.1111/emip.12293

Camara, W. J., & Westrick, P. (2017). *USA perspectives on admissions and admissions testing*. ACT. https://www.act.org/content/dam/act/unsecured/documents/R1653-usa-admission-perspectives-2017–09.pdf

Chen, X., & Simone, S. (2016). *Remedial coursetaking at U.S. public 2- and 4-year institutions: Scope, experiences, and outcomes* (Report No. 2016-405). U.S. Department of Education, Institute for Education Sciences, National Center for Education Statistics. https://nces.ed.gov/pubs2016/2016405.pdf

Chetty, R., Deming, D. J., & Friedman, J. N. (2023). *Diversifying society's leaders? The determinants and causal effects of admission to highly selective private colleges* [Working paper]. National Bureau of Economic Research. https://www.nber.org/papers/w31492

Clinedinst, M. (2019). 2019 *state of college admission*. National Association for College Admission Counseling. https://nacacnet.org/wp-content/uploads/2022/10/soca2019_all.pdf

College Board. (2016, May 9). *Concordance tables for new SAT scores*. https://satsuite.collegeboard.org/media/pdf/higher-ed-brief-sat-concordance.pdf

College Board. (2017a). *SAT suite of assessments technical manual: Characteristics of the SAT*. https://collegereadiness.collegeboard.org/pdf/sat-suite-assessments-technical-manual.pdf

College Board. (2017b). *SAT suite of assessments technical manual appendixes*. https://collegereadiness.collegeboard.org/pdf/sat-suite-assessments-technical-manual-appendix-pt-1.pdf

College Board. (2017c). *Next-generation ACCUPLACER: Test specifications (version 2.0)*. https://accuplacer.collegeboard.org/accuplacer/pdf/next-generation-test-specifications-manual.pdf

College Board. (2018). *ACCUPLACER*: *Program manual*. https://accuplacer.collegeboard.org/accuplacer/pdf/accuplacer-program-manual.pdf

College Board. (2019a). *SAT suite of assessments*. https://reports.collegeboard.org/sat-suite-program-results/class-2019-results

College Board. (2019b). *ACCUPLACER suite of assessments technical manual*.

College Board. (2024a, August). *Assessment framework for the digital SAT Suite* (Version 3.01). https://satsuite.collegeboard.org/media/pdf/assessment-framework-for-digital-sat-suite.pdf

College Board (2024b, March 12). *Digital SAT launches across the country, completing the transition to digital and providing a simpler testing experience for students and educators.* https://newsroom.collegeboard.org/digital-sat-launches-across-country-completing-transition-digital-and-providing-simpler-testing#:~:text=03/12/2024-,Digital%20SAT%20Launches%20Across%20the%20Country%2C%20Completing%20the%20Transition%20to,Board's%20new%20digital%20testing%20app

College Board. (2024c). *Digital SAT suite of assessments: Technical manual*. https://research.collegeboard.org/media/pdf/Digital%20SAT%20Suite%20of%20Assessments%20Technical%20Manual-FINAL.pdf

College Board & ACT. (2018). *Guide to the 2018 ACT/SAT concordance.* https://www.act.org/content/dam/act/unsecured/documents/ACT-SAT-Concordance-Information.pdf

Cruce, T. M., & Sanchez, E. I. (2023, April 12–15). *Score-sending to test-optional colleges and first-year college success* [Paper presentation]. National Council on Measurement in Education Annual Meeting, Chicago, IL, United States.

Dalessandro, S. P., Anthony, L. C., & Reese, L. M. (2014). *LSAT performance with regional, gender, and racial/ethnic breakdowns: 2007–2008 through 2013–2014 testing years* (LSAT Technical Report No. TR-14-02). Law School Admission Council.

Davenport, E. K., Howard, B., & Harrington Weston, S. (2018). An examination of the effect of the University of Michigan cases on the complexion of higher education. *Alabama Journal of Educational Leadership, 5*, 29–38. https://files.eric.ed.gov/fulltext/EJ1194727.pdf

Davies, M. (2011). Introduction to the special issue on critical thinking in higher education. *Higher Education Research & Development, 30*(3), 255–260. https://doi.org/10.1080/07294360.2011.562145

de Brey, C., Zhang, A., & Duffy, S. (2022). *Digest of education statistics, 2020* (NCES 2022-009). National Center for Education Statistics, Institute of Education Sciences, U.S. Department of Education. https://nces.ed.gov/programs/digest/d20/tables/dt20_305.40.asp

Donlon, T. (1984). *College Board technical handbook for the Scholastic Aptitude Test and achievement tests*. College Entrance Examination Board Publishing.

Donnon, T., Paolucci, E. O., & Violato, C. (2007). The predictive validity of the MCAT for medical school performance and medical board licensing examinations: A meta-analysis of the published research. *Academic Medicine, 82*, 100–106. https://doi.org/10.1097/01.acm.0000249878.25186.b7

Dorans, N. J., Lyu, C. F., Pommerich, M., & Houston, W. M. (1997). Concordance between ACT assessment and recentered SAT I sum scores. *College and University, 73*(2), 24–32.

Dudley, R. (2016, December 12). *College Board faces rocky path after CEO pushes new vision for SAT*. Reuters Investigates. https://www.reuters.com/investigates/special-report/college-sat-coleman/.

Dunleavy, D. M., Kroopnick, M. H., Dowd, K. W., Searcy, C. A., & Zhao, X. (2013). The predictive validity of the MCAT exam in relation to academic performance through medical school: A national cohort study of 2001–2004 matriculates. *Academic Medicine, 88*(5), 666–671. https://doi.org/10.1097/acm.0b013e3182864299

The Economist. (2014, April 5). *Is college worth it?* http://www.economist.com/news/united-states/21600131-too-many-degrees-are-waste-money-return-higher-education-would-be-much-better

Education Commission of the States. (2022, December). *50-State comparisons: States' school accountability*. https://www.ecs.org/50-state-comparison-states-school-accountability-systems/

Education Week. (2020, April 3). *State-by-state map of school-building closure*. https://www.edweek.org/ew/section/multimedia/map-coronavirus-and-school-closures.html

Edwards, A., Goyer, P., Howell, J., Hurwitz, M., Imlay, S., & Perfetto, G. (2023). *College outcomes following pandemic-induced changes in college admissions policies*. College Board. https://research.collegeboard.org/media/pdf/ARC_Research_Brief_2_04032023.pdf

ETS. (2017). *GRE to GMAT comparison tool for business schools*. https://www.ets.org/content/dam/ets-org/pdfs/gre/gre-bschool-comparison-tool.xlsx

ETS. (2020a). *Taking the GRE general test for law school*. https://www.ets.org/gre/revised_general/about/law/

ETS. (2020b). *The GRE® General Test*. https://www.ets.org/gre/institutions/about/general?WT.ac=rx01

ETS. (2022). *GRE General Test preparation guide*. Retrieved from https://www.ets.org/pdfs/gre/gre-guide-to-the-use-of-scores.pdf

ETS. (2023a). *GRE comparison tool for law schools*. https://www.ets.org/gre/score-users/law-schools/comparison-tool.html

ETS. (2023b). *Learn about GRE*. https://www.ets.org/gre/test-takers/general-test/about.html

ETS. (2023c). *The GRE Tests*. https://www.ets.org/gre/score-users/about/general-test/content-structure.html

ETS. (2024). *GRE guide to the use of scores*. https://www.ets.org/pdfs/gre/gre-guide-to-the-use-of-scores.pdf

FairTest (2024). *Test optional and test free colleges*. https://fairtest.org/test-optional-list/

Fein, P. (2015, June 18). *Finding a new compass*. Inside Higher Education. https://www.insidehighered.com/news/2015/06/18/act-drops-popular-compass-placement-test-acknowledging-its-predictive-limits

Feldon, D. F., Litson, K., Cahoon, B., Feng, Z., Walker, A., & Tofel-Grehl, C. (2023). The predictive validity of the GRE across graduate outcomes: A meta-analysis of trends over time. *Journal of Higher Education 95*(1), 120–148. https://doi.org/10.1080/00221546.2023.2187177

Felton, E. (2015, October 29). How the Common Core is transforming the SAT. *The Atlantic*. https://www.theatlantic.com/education/archive/2015/10/the-common-core-sat/412960/

Fields, R., & Parsad, B. (2012). *Tests and cut scores used for student placement in postsecondary education: Fall 2011*. National Assessment Governing Board. https://www.nagb.gov/content/dam/nagb/en/documents/what-we-do/preparedness-research/surveys/test-and-cut-scores-used-for-student-placement-in-postsecondary-education-fall-2011.pdf

Finn, B. (2015). *Measuring motivation in low-stakes assessments* (ETS Research Report No. RR-15-19). ETS. https://onlinelibrary.wiley.com/doi/10.1002/ets2.12067

Fisher v. University of Texas et al., 579 U.S. 136 (2016). https://www.oyez.org/cases/2015/14-981

Fuess, C. M. (1950). *College Board: Its first fifty years*. Columbia University Press.

Furuta, J. (2017). Rationalization and student/school personhood in U.S. college admissions: The rise of test-optional policies, 1987 to 2015. *Sociology of Education*, *90*(3), 236–254. https://doi.org/10.1177/0038040717713583

Gaff, J., & Meachum, J. (2006). Learning goals in mission statements: Implications for educational leadership. *Liberal Education*, *92*(1), 6–13. https://files.eric.ed.gov/fulltext/EJ743272.pdf

Gneezy, U., List, J. A., Livingston, J. A., Sadoff, S., Qin, X., & Xu, Y. (2017). *Measuring success in education: The role of effort on the test itself* (NBER Working Paper 24004). National Bureau of Economic Research.

Goldberg, J., & Smith, J. (2008). The effects of education on labor market outcomes. In H. F. Ladd & E. B. Fiske (Eds.), *Handbook of research in education finance and policy* (pp. 688–708). Routledge.

Graduate Management Admission Council. (2011). *Next-Generation GMAT exam*. https://www.gmac.com/gmat/next-generation-gmat-information-center/~/media/files/gmac/gmat/2011_nextgenerationgmat_exam_v3.pdf

Graduate Management Admission Council. (2018). *GMAT exam user guide for institutions*. https://www.gmac.com/-/media/files/gmac/gmat/gmat_institutionuserguide_fnl.pdf

Graduate Management Admission Council. (2020). *GMAC history*. https://www.gmac.com/why-gmac/the-gmac-advantage/gmac-history

Graduate Management Admission Council. (2023). *Exam structure, content & features*. https://www.gmac.com/gmat-other-assessments/about-the-gmat-focus-edition/exam-overview

Graduate Management Admission Council. (2024a). *The GMAT Exam: Looking forward*. https://www.gmac.com/gmat-other-assessments/about-the-gmat-focus-edition/gmat-focus-edition

Graduate Management Admission Council (2024b). *Understanding your score*. https://www.mba.com/exams/gmat-exam/scores/understanding-your-score

Graduate Management Admission Council. (2025). *Scores*. https://www.mba.com/exams/gmat-exam/scores

Gratz v. Bollinger et al., 539 U.S. 244 (2003). https://www.oyez.org/cases/2002/02-516

Griffith, R. L., Wolfeld, L., Armon, B. K., Rios, J., & Liu, O. L. (2016). *Assessing intercultural competence in higher education: Existing research and future directions* (ETS Research Report No. RR-16-25). ETS. https://onlinelibrary.wiley.com/doi/epdf/10.1002/ets2.12112

Grutter v. Bollinger et al., 539 U.S. 306 (2003). https://www.oyez.org/cases/2002/02-241

Han, C. (2022, October 18). *GRE to GMAT conversion: Why you can't compare scores*. MBA.com. https://www.mba.com/exams-and-exam-prep/gmat-exam/gmat-gre-why-you-cant-compare-scores

Han, K. T. (2025, October 17). *GMAT Focus psychometric updates* [Seminar presentation]. GMAT Test Preparation Summit, Reston, VA, United States.

Hartocollis, A., & Saul, S. (2024, April 11). Harvard and Caltech will require test scores. *New York Times*. https://www.nytimes.com/2024/04/11/us/harvard-test-scores-admissions.html

Hawkins, D. (2019). Operation Varsity Blues: Where we are now. *Journal of College Admission, 244*, 15–17. https://eds.p.ebscohost.com/eds/pdfviewer/pdfviewer?vid=0&sid=b909a86e-6c63-4afd-8b10-1966b2c524b8%40redis

Hiss, W., & Franks, V. (2014). *Defining promise: Optional standardized testing policies in American college and university admissions*. Lumina Foundation. https://www.luminafoundation.org/files/resources/definingpromise.pdf

Howell, J. (2023, April 12–15). *Insights from the admissions research consortium* [Paper presentation]. National Council on Measurement in Education Annual Meeting, Chicago, IL, United States.

Hoxby, C., & Avery, C. (2013). *The missing "one-offs": The hidden supply of high-achieving, low-income students*. Brookings Papers on Economic Activity. https://www.brookings.edu/wp-content/uploads/2016/07/2013a_hoxby.pdf

Hurwitz, M., & Lee, J. (2018). Grade inflation and the role of standardized testing. In J. Buckley, L. Letukas, & B. Wildavsky (Eds.), *Measuring success: Testing, grades and the future of college admissions* (pp. 64–93). Johns Hopkins University Press.

Hyman, J. (2017). ACT for all: The effect of mandatory college entrance exams on postsecondary attainment and choice. *Education Finance and Policy, 12*(3), 281–311. https://doi.org/10.1162/edfp_a_00206

Inside Higher Ed. (2019). *2019 survey of college and university presidents*. https://www.insidehighered.com/news/survey/2019-survey-college-and-university-presidents

Jacobsen, E. (2017). *A (mostly) brief history of the SAT and ACT tests*. Erik the Red. https://erikthered.com/tutor/sat-act-history.html

Jankowski, N. A., Timmer, J. D., Kinzie, J., & Kuh, G. D. (2018, January). *Assessment that matters: Trending toward practices that document authentic student learning*. University of Illinois and Indiana University, National Institute for Learning Outcomes Assessment. https://files.eric.ed.gov/fulltext/ED590514.pdf

Jaschik, S. (2017, September 18). *At University of Texas at Austin, 10% has become 6%*. Inside Higher Ed. https://www.insidehighered.com/admissions/article/2017/09/18/university-texas-austin-adjusts-downward-percent-plan-used-admissions

Jaschik, S. (2018, April 9). *GMAT loses 30 minutes*. Inside Higher Ed. https://www.insidehighered.com/admissions/article/2018/04/09/gmat-facing-competition-gre-cuts-time-test-30-minutes

Jaschik, S. (2020, October 19). *ACT and SAT scores drop*. Inside Higher Ed. https://www.insidehighered.com/admissions/article/2020/10/19/act-and-sat-scores-drop-2020?msclkid=994ac0f4a60211eca79677b819f8b0b0

Jencks, C., & Phillips, M. (Eds.). (1998). *The Black–White test score gap*. Brookings Institution Press.

Julian, E. R. (2005). Validity of the Medical College Admission Test for predicting medical school performance. *Academic Medicine, 80*(10), 910–917. https://doi.org/10.1097/00001888-200510000-00010

Julian, T. (2012). *Work–life earnings by field of degree and occupation for people with a bachelor's degree: 2011* (Report No. ACSBR/11-04). U.S. Census Bureau. https://www.census.gov/library/publications/2012/acs/acsbr11-04.html

Kaniuka, T., & Wynne, M. (2019). Exploring the relationship between the Collegiate Learning Assessment, student learning activities, and study behaviors: Implications for colleges and universities. *Education Research International, 2019*(1), 1–11. https://doi.org/10.1155/2019/3648318

Kaplan. (2021, February 9). *Kaplan survey: Nearly 30 percent of colleges say the Black Lives Matter movement changed their admissions process* [Press release]. https://

www.kaptest.com/blog/press/2021/02/09/kaplan-survey-nearly-30-percent-of-colleges-say-the-black-lives-matter-movement-changed-their-admissions-process/

Keiser, H. N., Sackett, P. R., Kuncel, N. R., & Brothen, T. (2016). Why women perform better in college than admission scores would predict: Exploring the roles of conscientiousness and course-taking patterns. *Journal of Applied Psychology*, *101*(4), 569–581. https://doi.org/10.1037/apl0000069

Kelly, R., & Morgan, J. (2024). LSAT performance with regional, gender, racial and ethnic, repeater, and disability breakdowns: 2018-2019 through 2022-2023 testing years (LSAT Technical Report Series TR 24-01). LSAC. https://www.lsac.org/sites/default/files/research/TR-24-01.pdf

Kent, J. D., & McCarthy, M. T. (2016). *Holistic review in graduate admissions: A report from the Council of Graduate Schools*. Council of Graduate Schools.

Klasik, D. (2013). The ACT of enrollment: The college enrollment effects of state-required college entrance exam testing. *Educational Researcher*, *42*(3), 151–160. https://doi.org/10.3102/0013189x12474065

Klein, S., Liu, O. L., Sconing, J., Bolus, R., Bridgeman, B., Kugelmass, H., Nemeth, A., Robbins, S., & Steedle, J. (2009). *Test validity study (TVS) report.* Collegiate Learning Assessment.

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* (ETS Research Report No. RR-18-26). ETS. https://doi.org/10.1002/ets2.12213

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* (ETS Research Report No. RR-14-26). ETS. http://dx.doi.org/10.1002/ets2.12026

Knox, L. (2023, October, 19). No more "logic games" on the LSAT. *Inside Higher Ed.* https://www.insidehighered.com/news/quick-takes/2023/10/19/lsat-drops-analytical-reasoning-section-after-lawsuit

Knox, L. (2024). Introducing a shorter, 'more flexible' ACT. *Inside Higher Ed*. https://www.insidehighered.com/news/admissions/traditional-age/2024/07/22/act-announces-shorter-cheaper-science-optional-exam

Kobrin, J. L., Patterson, B. F., Shaw, E. J., Mattern, K. D., & Barbuti, S. (2008). *Validity of the SAT for predicting first-year college grade point average* (Report No. 2008-5). College Board. https://files.eric.ed.gov/fulltext/ED563202.pdf

Kuncel, N. R., Credé, M., & Thomas, L. L. (2007). A meta-analysis of the predictive validity of the Graduate Management Admission Test (GMAT) and undergraduate grade point average (UGPA) for graduate student academic performance. *Academy of Management Learning & Education*, *6*(1), 51–68. https://doi.org/10.5465/amle.2007.24401702

Kuncel, N. R., & Hezlett, S. A. (2007). Standardized tests predict graduate students' success. *Science*, *315*(5815), 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*(1), 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 Examination for master's and doctoral programs: A meta-analytic investigation. *Educational and Psychological Measurement, 70,* 340–352. https://doi.org/10.1177/0013164409344508

Laird, R. (2005). What is it we think we are trying to fix and how should we fix it? A view from the admissions office. In W. J. Camara & E. W. Kimmel (Eds.), *Choosing students: Higher education admissions tools for the 21st century* (pp. 13–32). Routledge.

LaPrade, A., Mertens, J., Moore, T., & Wright, A. (2019). *The enterprise guide to closing skills gap, strategies for building and maintaining a skilled workforce*. Institute for Business Value. https://www.ibm.com/thought-leadership/institute-business-value/report/closing-skills-gap

Lawrence, I., Rigol, G. W., Van Essen, T., & Jackson, C. A. (2002). *A historical perspective on the SAT 1926-2001* (Report No. 2002-7). College Board.

Law School Admission Council. (2019). *How our test contributes to comprehensive and fair law school admissions*. https://www.lsac.org/blog/how-our-test-contributes-comprehensive-and-fair-law-school-admissions.

Law School Admission Council. (2020). *LSAT trends: Total LSATs administered by admin & year*. https://report.lsac.org/TestTakers.aspx?Format=PDF

Law School Admission Council. (2022). *2021–22 interpretive guide for LSAT® score users*. LSAC.

Law School Admission Council. (2023a). *Summary of 2017, 2018, and 2019 LSAT correlation study results*. (2nd ed.) https://www.lsac.org/data-research/research/summary-2017-2018-and-2019-lsat-correlation-study-results

Law School Admission Council. (2023b). *What to expect starting with the August 2024 LSAT*. https://www.lsac.org/blog/what-to-expect-starting-with-august-2024-lsat

Law School Admission Council. (2024, December 18.) *Admission trends: Applicants, admitted applicants, matriculants, & applications*. https://report.lsac.org/View.aspx?Report=AdmissionTrendsApplicantsAdmitApps&Format=PDF

Lawyers for Civil Rights v. President and Fellows of Harvard College. Complaint before U.S. Department of Education (2023, July 3).

Lemann, N. (1999). *The big test: The secret history of the American meritocracy*. Farrar, Straus and Giroux.

Lewin, T. (2009, December 9). A new look for graduate entrance test. *The New York Times*. https://www.nytimes.com/2009/12/06/education/06gre.html

Li, D., Yi, Q., & Harris, D. (2017). *Evidence for paper and online ACT comparability: Spring 2014 and 2015 mode comparability studies* (Report No. 2017-1). ACT.

Lindquist, E. F. (1959). The nature of the problem of improving scholarship and college entrance examinations. In *Proceedings of the Invitational Conference on Testing Problems* (pp. 104–113). ETS.

Liu, O. L. (2017). Ten years after the Spellings Commission: From accountability to internal improvement. *Educational Measurement: Issues and Practice, 36*(2), 34–41. http://dx.doi.org/10.1111/emip.12139.

Liu, O. L. (2023). Assessing student learning outcomes of higher education. In C. Baik & E. Kahu (Eds.), *Research handbook on the student experience in higher education* (pp. 283–295). Edward Elgar Publishing.

Liu, O. L., Bridgeman, B., & Adler, R. M. (2012). Measuring learning outcomes assessment in higher education: Motivation matters. *Educational Researcher, 41*, 352–362. https://psycnet.apa.org/doi/10.3102/0013189X12459679

Liu, O. L., Bridgeman, B., Gu, L. X., Xun, J., & Kong, N. (2015). Investigation of response change on the revised GRE. *Educational & Psychological Measurement, 75*(6), 1002–1020. https://dx.doi.org/10.1177%2F0013164415573988

Liu, O. L., Frankel, L., & Roohr, K. C. (2014). *Assessing critical thinking in higher education: Current state and directions for next generation assessment* (ETS Research Report No. RR-14-10). ETS. https://doi.org/10.1002/ets2.12009

Liu, O. L., Kell, H. J., Liu, L., Ling, G., Wang, Y., Wylie, C., Sevak, A., Sherer, D., LeMahieu, P., & Knowles, T. (2023). *A new vision for skills-based assessment.* ETS. https://www.ets.org/pdfs/rd/new-vision-skills-based-assessment.pdf

Liu, O. L., Kell, H., Williams, K., Ling, G. M., & Sanders, M. (2023). ETS 2025 skills taxonomy. *Chinese/English Journal of Educational Measurement and Evaluation. 4*(4), Article 2. https://doi.org/10.59863/URZL9477

Liu, O. L., Klieger, D. M., Bochenek, J. L., Holtzman, S. L., & Xu, J. (2016). *An investigation of the use and predictive validity of scores from the GRE® revised General Test in a Singaporean university* (ETS Research Report No. RR-16-05). ETS. https://doi.org/10.1002/ets2.12095

Liu, O. L., Liu, H., Roohr, K. C., & McCaffrey, D. F. (2016). Investigating college learning gain: Exploring a propensity score weighting approach. *Journal of Educational Measurement, 53*(3), 352–367. https://doi.org/10.1111/jedm.12112

Liu, O.L., Ling, G., & Fishtein, D. (2021). Charting the skills for the future: Perspectives from employers and higher education institutions. *NCME Chinese/English Journal of Educational Measurement and Evaluation 3*(1), 24–32.

Lucido, J. A. (2018). Understanding the test-optional movement. In J. Buckley, L. Letukas, & B. Wildavsky (Eds.), *Measuring success: Testing, grades and the future of college admissions* (pp. 146–170). Johns Hopkins University Press.

Magouirk, P., Freeman, M., Kajkawa, T., Karimi, H., & Kim, B. H. (2023, January 11). *Deadline update: first-year application trends through January 1*. Common App. https://s3.us-west-2.amazonaws.com/ca.research.publish/Deadline+Updates/DeadlineUpdate_010323.pdf

Marco, G. L., Abdel-fattah, A. A., & Baron, P. A. (1992). *Methods used to establish score comparability on the enhanced ACT assessment and the SAT* (Research Report No. 92-3). College Board. https://doi.org/10.1002/J.2333-8504.1992.TB01456.X

Margolis, W. (2018, December 10). *LSAC announces technology collaboration with Microsoft*. Law School Admission Council. https://www.lsac.org/about/news/lsac-announces-technology-collaboration-microsoft

Marini, J. P., Westrick, P. A., Young, L., Ng, H., & Shaw, E. J. (2023). *Digital SAT pilot predictive validity study—A first look*. College Board.

Marini, J. P., Westrick, P. A., Young, L., Ng, H., Shmueli, D., & Shaw, E. J. (2019). *Differential validity and prediction of the SAT®: Examining first-year grades and retention to the second year*. College Board. https://files.eric.ed.gov/fulltext/ED597325.pdf

Marion, S., & Domaleski, C. (2019). An argument in search of evidence: A response to Camara et al. (2019). *Educational Measurement: Issues and Practice, 38*(4), 27–28. http://doi.org/10.1111/emip.12307

Mattern, K., & Allen, J. (2016). *More information, more informed decisions: Why test-optional policies do not benefit institutions or students*. ACT. https://www.act.org/content/dam/act/unsecured/documents/R1603_More-Information-More-Informed-Decisions_Web.pdf

Mattern, K. D., & Packman, S. (2009). *Predictive validity of ACCUPLACER scores for course placement: A meta-analysis* (Research Report No. 2009-2). College Board. https://files.eric.ed.gov/fulltext/ED561046.pdf

Mattern, K. D., & Patterson, B. F. (2014). *Synthesis of recent SAT validity findings: Trend data over time and cohorts* (College Board Research in Review 2014-1). College Board. https://files.eric.ed.gov/fulltext/ED556462.pdf

Mattern, K. D., Patterson, B. F., Shaw, E. J., Kobrin, J. L., & Barbuti, S. (2008). *Differential validity and prediction of the SAT* (Research Report No. 2008-4). College Board. https://files.eric.ed.gov/fulltext/ED562614.pdf

Mattern, K., Radunzel, J., Bertling, M., & Ho, A. D. (2018). How should colleges treat multiple admissions test scores? *Educational Measurement: Issues and Practice, 37*(3), 11–23. https://doi.org/10.1111/emip.12199

Mattern, K., Sanchez, E., & Ndum, E. (2017). Why do achievement measures underpredict female academic performance? *Educational Measurement: Issues and Practice, 36*(1), 47–57. https://psycnet.apa.org/doi/10.1111/emip.12138

Mattern, K., & Steedle, J. (2020). *Students who take the ACT test sections in a different order earn similar scores* [ACT Research Report 2020-1]. ACT. https://www.act.org/content/dam/act/unsecured/documents/R1822-section-order-10-2020.pdf

McGaghie, W. (2002). Assessing readiness for medical education: Evolution of the Medical College Admission Test. *Journal of the American Medical Association, 288*(9), 1085–1090. https://doi.org/10.1001/jama.288.9.1085

McKinsey Global Institute. (2017, December). *Jobs lost, jobs gained: Workforce transitions in a time of automation*. McKinsey & Company. https://www.mckinsey.com/~/media/mckinsey/featured%20insights/Future%20of%20Organizations/What%20the%20future%20of%20work%20will%20mean%20for%20jobs%20skills%20and%20wages/MGI-Jobs-Lost-Jobs-Gained-Report-December-6-2017.ashx

Michel, R. S., Belur, V., Naemi, B., & Kell, H. J. (2019). *Graduate admissions practices: A targeted review of the literature* (ETS Research Report No. RR-19-33). ETS. https://onlinelibrary.wiley.com/doi/pdf/10.1002/ets2.12271

Monroe, A., Quinn, E., Samuelson, W., Dunleavy, D. M., & Dowd, K. W. (2013). An overview of the medical school admission process and use of applicant data in

decision making: What has changed since the 1980s? *Academic Medicine*, *88*(5), 672–281. https://doi.org/10.1097/acm.0b013e31828bf252

National Association of Colleges and Employers. (2018). *Job outlook 2019*. https://www.odu.edu/content/dam/odu/offices/cmc/docs/nace/2019-nace-job-outlook-survey.pdf

National Center for Education Statistics. (2018). *Digest of education statistics, Table 302.10. Recent high school completers and their enrollment in college, by sex and level of institution: 1960 through 2017*. U.S. Department of Education, Institute of Education Sciences. https://nces.ed.gov/programs/digest/d18/tables/dt18_302.10.asp

National Council on Measurement in Education. (2019). *National Council on Measurement in Education position statement on the use of college admissions test scores as academic indicators in state accountability systems*. https://doi.org/10.1111/emip.12309

National Institute for Learning Outcomes Assessment. (2019, September). *Student learning outcomes statements*. https://www.learningoutcomesassessment.org/ourwork/transparency-framework/components/student-learning-outcomes-statements/

National Student Clearinghouse Research Center. (2019). *Snapshot report: First-year persistence and retention for fall 2017 cohort*. https://nscresearchcenter.org/wp-content/uploads/SnapshotReport35.pdf

Nunnally, J. C. (1978). *Psychometric theory* (2nd ed.). McGraw–Hill.

Office of the President of University of California. (2020, May 21). *Executive summary: College entrance exam use in University of California undergraduate admissions*. https://regents.universityofcalifornia.edu/regmeet/may20/b4.pdf

Oh, I. S., Schmidt, F. L., Shaffer, J. A., & Le, H. (2008). The Graduate Management Admission Test (GMAT) is even more valid than we thought: A new development in meta-analysis and its implications for the validity of the GMAT. *Academy of Management Learning & Education*, *7*(4), 563–570. https://doi.org/10.5465/amle.2008.35882196

Olson, E. (2017, August 10). More law schools begin accepting GRE test results. *New York Times*. https://www.nytimes.com/2017/08/10/business/dealbook/law-school-gre.html

Organisation for Economic Co-operation and Development. (2013). *Assessment of higher education learning outcomes: Feasibility study report*: *Vol. 3. Further insights*. http://www.oecd.org/education/skills-beyond-school/AHELOFSReportVolume3.pdf

Pellegrino, J. W., & Hilton, M. L. (2012). *Education for life and work: Developing transferable knowledge and skills in the 21st century*. National Research Council, The National Academies Press. https://www.nap.edu/read/13398/chapter/1

Posselt, J. R. (2016). *Inside graduate admissions: Merit, diversity, and faculty gatekeeping*. Harvard University Press.

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*, 209–219. https://doi.org/10.1037/0021-9010.89.2.208

Powers, D. E., & Fowles, M. E. (2000). *Likely impact of the GRE writing assessment on graduate admissions decisions* (ETS Research Report No. RR-00-16). ETS. http://dx.doi.org/10.1002/j.2333-8504.2000.tb01839.x

Radunzel, J., & Mattern, K. (2020a). *Predicting students' chances of completing a degree: How does superscoring compare to other scoring methods when applicants retest?* (Working Paper 2020-01). ACT. https://www.act.org/content/dam/act/unsecured/documents/R1825-degree-completion-superscore-2020-05.pdf

Radunzel, J., & Mattern, K. (2020b). *A case study: ACT section retest scores and superscores are predictive of first-term grades*. ACT. https://files.eric.ed.gov/fulltext/ED606153.pdf

Radunzel, J., Mattern, K., Crouse, J., & Westrick, P. (2015). *Development and validation of a STEM benchmark based on the ACT STEM score*. ACT. https://www.act.org/content/dam/act/unsecured/documents/2015-Tech-Brief-Development-and-Validation.pdf

Radunzel, J., & Noble, J. (2012). *Predicting long-term college success through degree completion using ACT Composite score, ACT Benchmarks, and high school grade point average* (Report No. 2012-5). ACT. https://www.act.org/content/dam/act/unsecured/documents/ACT_RR2012-5.pdf

Radunzel, J., & Noble, J. (2013). *Differential effects on student demographic groups of using ACT college readiness assessment composite score, ACT benchmarks, and high school grade point average for predicting long-term college success through degree completion* (Report No. 2013-5). ACT.

Rahman, N., & Trierweiler, T. J. (2017). *Analysis of differential prediction of law school performance by race/ethnicity based on 2011–2014 entering law school classes* (LSAT Technical Report No. TR-17-01). Law School Admission Council.

Reese, L. M., & Cotter, R. A. (1994). *A compendium of LSAT and LSAC-sponsored item types 1948–1994* (LSAT Research Report No. RR-94-01). Law School Admission Council. https://files.eric.ed.gov/fulltext/ED469242.pdf

Richer, A. D., & Binkley, C. (2019, March 12). *TV stars and coaches charged college bribery scheme*. Associated Press. https://apnews.com/2450688f9e67435c8590e59a1b0e5b47

Rigol, G. (2003). *Admissions decision-making models: How US institutions of higher education select undergraduate students*. College Entrance Examination Board. https://files.eric.ed.gov/fulltext/ED562589.pdf

Rigol, G. (2004). *Selection through individual review: A report on phase IV of the admissions models project*. College Entrance Examination Board. https://files.eric.ed.gov/fulltext/ED562855.pdf

Rippetoe, R. (2021, June 20). *Winston & Strawn aids UC students in test score case*. Law360. https://www.law360.com/articles/1394856/winston-strawn-aids-uc-students-in-test-score-case

Rodriguez, O., Bowden, B., Belfield, C., & Scott-Clayton, S. (2015). *Calculating the costs of remedial placement testing*. Community College Research Center. https://ccrc.tc.columbia.edu/media/k2/attachments/calculating-cost-remedial-placement-analytics-2.pdf

Roohr, K. C., Graf, E. A., & Liu, O. L. (2014). *Assessing quantitative literacy in higher education: An overview of existing research and assessments with recommendations for next generation assessment* (ETS Research Report No. RR-14-22). ETS. https://doi.org/10.1002/ets2.12024

Roohr, K., Olivera-Aguilar, M., Ling, G., & Rikoon, S. (2019). A multi-level modeling approach to investigating students' critical thinking at higher education institutions. *Assessment & Evaluation in Higher Education, 44*(2), 1–15. https://doi.org/10.1080/02602938.2018.1556776

Rutschow, E. Z., & Mayer, A. K. (2018). *Early findings from a national survey of developmental education practices* [Research brief]. Center for the Analysis of Postsecondary Readiness. https://www.mdrc.org/sites/default/files/2018_CAPR_Descriptive_Study.pdf

Sanchez, E. I. (2013). *Differential effects of using ACT college readiness assessment scores and high school GPA to predict first-year college GPA among racial/ethnic, gender, and income groups* (Report No. 2013-4). ACT. https://www.act.org/content/dam/act/unsecured/documents/ACT_RR2013-4.pdf

Sawyer, R. (2010). *Usefulness of high school average and ACT scores in making college admission decisions* (Report No. 2010-2). ACT. https://www.act.org/content/dam/act/unsecured/documents/ACT_RR2010-2.pdf

Sawyer, R. (2013). Beyond correlations: Usefulness of high school GPA and test scores in making college admissions decisions. *Applied Measurement in Education, 26*(2), 89–112. https://doi.org/10.1080/08957347.2013.765433

Schemo, D. J. (2001, February 17). Head of U. of California seeks to end SAT use in admissions. *New York Times*. https://www.nytimes.com/2001/02/17/us/head-of-u-of-california-seeks-to-end-sat-use-in-admissions.html

Schudson, M. S. (1972). Organizing the "meritocracy": A history of the college entrance examination board. *Harvard Educational Review, 42*(1), 34–69. https://doi.org/10.17763/haer.42.1.l5772104r6108v01

Schwager, I. T., Hülsheger, U. R., Bridgeman, B., & Lang, J. W. (2015). Graduate student selection: Graduate Record Examination, socioeconomic status, and undergraduate grade point average as predictors of study success in a western European University. *International Journal of Selection and Assessment, 23*(1), 71–79. https://doi.org/10.1111/ijsa.12096

Scott-Clayton, J. (2012). *Do high-stakes placement exams predict college success?* (CCRC Working Paper No. 41). Columbia University, Teachers College, Community College Research Center. http://ccrc.tc.columbia.edu/media/k2/attachments/high-stakes-predict-success.pdf

Shapiro, D., Dundar, A., Huie, F., Wakhungu, P. K., Bhimdiwali, A., Nathan, A., & Youngsik, H. (2018). *Transfer and mobility: A national view of student movement in postsecondary institutions, fall 2011 cohort* (Signature Report No. 15). National Student Clearinghouse.

Shaw, E. J., Marini, J. P., Beard, J., Shmueli, D., Young, L., & Ng, H. (2016). *The redesigned SAT® pilot predictive validity study: A first look* (Report No. 2016-1). College Board.

Shelton, P. D. (1997). The LSAT: Good—but not that good. *Law School Services Report*, *97*, 2–3.

Shultz, M. M., & Zedeck, S. (2011). Predicting lawyer effectiveness: Broadening the basis for law school admission decisions. *Law & Social Inquiry*, *36*(3), 620–661. https://doi.org/10.1111/j.1747-4469.2011.01245.x

Smith v. Regents of University of California et al. (2020). https://unicourt.com/case/ca-sca1-smith-v-regents-of-the-university-of-california-et-al-131183

Snyder, T. D., de Brey, C., & Dillow, S. A. (2019). *Digest of Education Statistics 2017* (NCES 2018-070; Table 219.10, p. 193). National Center for Education Statistics, Institute of Education Sciences, U.S. Department of Education. https://nces.ed.gov/pubs2018/2018070.pdf

Soares, J. A. (2012). *SAT wars: The case for test-optional college admissions*. Teachers College.

Sparks, J. R., Song, Y., Brantley, W., & Liu, O. L. (2014). *Assessing written communication in higher education: Review and recommendations for next generation assessment* (ETS Research Report No. RR-14-37). ETS. https://doi.org/10.1002/ets2.12035

Standardized Testing Task Force. (2020, January). *Report of the UC Academic Council Standardized Testing Task Force (UC Task Force)*. https://senate.universityofcalifornia.edu/_files/committees/sttf/sttf-report.pdf

Stilwell, L. A., Dalessandro, S. P., & Reese, L. M. (2011). *Predictive validity of the LSAT: A national summary of the 2009 and 2010 LSAT correlation studies* (LSAT Technical Report TR-11-02). Law School Admission Council. https://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.225.3802&rep=rep1&type=pdf

Stilwell, L. A., Diamond-Dalessandro, S. P., & Reese, L. M. (2007). *Predictive validity of the LSAT: A national summary of the 2005–2006 correlation studies* (LSAT Technical Report TR-07-02). Law School Admission Council.

Stilwell, L. A., Diamond-Dalessandro, S. P., & Reese, L. M. (2009). *Predictive validity of the LSAT: A national summary of the 2007–2008 correlation studies* (LSAT Technical Report TR-09-03). Law School Admission Council.

Students for Fair Admissions v. President and Fellows of Harvard College. (2023). Oyez. https://www.oyez.org/cases/2022/20-1199.

Students for Fair Admissions v. University of North Carolina. (2023). Oyez. https://www.oyez.org/cases/2022/21-707.

Sweitzer, K., Blaock, A., E., Sharma, D. B. (2018). The effect of going test-optional on diversity and admissions: A propensity score matching analysis. In J. Buckley, L. Letukas, & B. Wildavsky (Eds.), *Measuring success: Testing, grades and the future of college admissions* (pp. 288–307). Johns Hopkins University Press.

Syverson, S. T., Franks, V. W., & Hiss, W. C. (2018). *Defining access: How test-optional works*. NACAC. https://www.nacacnet.org/globalassets/documents/publications/research/defining-access-report-2018.pdf

Talento-Miller, E. (2017). *Differential validity and differential prediction of the GMAT Exam*. Graduate Management Admission Council. https://www.gmac.com/-/media/files/gmac/research/validity-and-testing/rr-17-01-differential-validity-talento-miller-web-release-2.pdf

Torney-Purta, J., Cabrera, J. C., Roohr, K. C., Liu, O. L., & Rios, J. A. (2015). *Assessing civic competency and engagement in higher education: Research background, frameworks, and directions for next-generation assessment* (ETS Research Report No. RR-15-34). ETS. http://dx.doi.org/10.1002/ets2.12081

Tremblay, K., Lalancette, D., & Roseveare, D. (2012). *Assessment of higher education learning outcomes (AHELO): Feasibility study report: Vol 1. Design and implementation*. OECD. www.oecd.org/edu/skills-beyond-school/AHELOFSReportVolume1.pdf

Trierweiler, T. J., & Rahman, N. (2017). *Analysis of differential prediction of law school performance by gender based on 2011–2014 entering law school classes* (LSAT Technical Report TR-17-02). Law School Admission Council. https://www.lsac.org/data-research/research/analysis-differential-prediction-law-school-performance-gender-based-2011

University of California, Ethics, Compliance, and Audit Services. (2020, February). *System wide audit of undergraduate admissions—Phase 2* (Project No. P20A005). https://www.universityofcalifornia.edu/sites/default/files/second-audit-report.pdf

U.S. Department of Education. (2006). *A test of leadership: Charting the future of U.S. higher education*. The Secretary of Education's Commission on the Future of Higher Ed. https://files.eric.ed.gov/fulltext/ED493504.pdf

U.S. Department of Education. (2015). *Every Student Succeeds Act (ESSA)*. https://www.ed.gov/essa?src=rn

Wagenaar, R. (Ed.). (2018). *Tuning–CALOHEE assessment reference frameworks for civil engineering teacher education history nursing physics*. University of Groningen. https://www.calohee.eu/wp-content/uploads/2018/11/4.1-Assessment-Reference-Frameworks-for-Civil-Engineering-Teacher-Education-History-Nursing-and-Physics-FINAL-v2.pdf

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

Westrick, P. A., Le, H., Robbins, S. B., Radunzel, J. M., & Schmidt, F. L. (2015). College performance and retention: A meta-analysis of the predictive validities of ACT scores, high school grades, and SES. *Educational Assessment, 20*(1), 23–45. http://dx.doi.org/10.1080/10627197.2015.997614

Westrick, P. A., Marini, J. P., Shmueli, D., Young, L., Shaw, E. J., & Ng, H. (2020). *Validity of SAT® for predicting first-semester, domain-specific grades*. College Board. https://files.eric.ed.gov/fulltext/ED603085.pdf

Westrick, P. A., Marini, J. P., Young, L., Ng, H., Shmueli, D., & Shaw, E. J. (2019). *Validity of the SAT for predicting first-year grades and retention to the second year*. College Board. https://satsuite.collegeboard.org/media/pdf/national-sat-validity-study.pdf

Westrick, P. A., Schmidt, F. L., Le, H., Robbins, S. B., & Radunzel, J. M. (2021). The road to retention passes through first-year academic performance: A meta-analytic path analysis of academic performance and persistence. *Educational Assessment, 26*(1), 35–51. https://doi.org/10.1080/10627197.2020.1848423

Whitney, D. R. (1989). Education admissions ad placement. In R. Linn (Ed.), *Educational measurement* (3rd ed., pp. 515–526). Macmillan Publishing Co., American Council on Education.

Willingham, W. W. (1985). *Success in college: The role of personal qualities and academic ability*. College Entrance Examination Board.

Willingham, W. W., & Cole, N. S. (1997). *Gender and fair assessment*. Lawrence Erlbaum Associates.

Wise, S. L., & DeMars, C. (2005). Low examinee effort in low-stakes assessment: Problems and potential solutions. *Educational Assessment, 10*(1), 1–17. http://dx.doi.org/10.1207/s15326977ea1001_1

Young, J. W. (2004). Differential validity and prediction: Race and sex differences in college admissions testing. In R. Zwick (Ed.), *Rethinking the SAT: The future of standardized testing in university admissions* (pp. 289–301). Routledge.

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* (ETS Research Report No. RR-14-17). ETS. https://files.eric.ed.gov/fulltext/EJ1109333.pdf

Zamudio-Suarez, F. (2022, July 12). *Is remedial ed holding students of color back?* Race on Campus. https://www.chronicle.com/newsletter/race-on-campus/2022-07-12

Zhou, E., & Gao, J. (2021). *Graduate enrollment and degrees: 2010 to 2020*. Council of Graduate Schools. https://cgsnet.org/wp-content/uploads/2022/02/CGS_GED20_Report_final_v2-2.pdf

Zimmermann, J., von Davier, A., Buhmann, J. M., & Heinimann, H. R. (2017). Validity of GRE General Test scores and TOEFL scores for graduate admission to a technical university in western Europe. *European Journal of Engineering Education, 43*, 144–165. https://doi.org/10.1080/03043797.2017.1343277

Zlatkin-Troitschanskais, O., Toepper, M., Pant, H. A., Lautenbach, C., & Kuhn, C. (Eds.). (2018). *Methodology of educational measurement and assessment: Assessment of learning outcomes in higher education, cross-national comparisons and perspectives*. Springer.

Zwick, R. (2002). *Fair game? The use of standardized admissions tests in higher education*. Routledge.

Zwick, R. (2006). Higher education admissions testing. In R. L. Brennan (Ed.), *Educational measurement* (4th ed., pp. 647–679). Praeger Publishers.

Zwick, R. (2013). *Disentangling the role of high school grades, SAT® scores, and SES in predicting college achievement.* ETS Research Report No. RR-13-09.

Zwick, R. (2017). *Who gets in? Strategies for fair and effective college admissions*. Harvard University Press.

Zwick, R. (2018). How do percent plans and other test-optional admissions programs affect the academic performance and diversity of the entering class? In J. Buckey, L. Letukas, & B. Wildavsky (Eds.), *Measuring success: Testing, grades and the future of college admissions* (pp. 239–259). Johns Hopkins University Press.

NOTES

1. Data were obtained through the National Center for Education Statistics Integrated Postsecondary Education Data System data center at https://nces.ed.gov/ipeds/datacenter/reportmain.aspx. Enrollment of all students at test-optional colleges represented about 5% of all freshmen nationally in 2018, but since about 70%–80% of students at such colleges submit test scores, the estimate is that students admitted without scores represent 1%–1.25% of freshmen.
2. See page 1194 in this chapter for a discussion of *Smith v. Regents of University of California*, 2020.
3. See Rigol (2004) for an example of an admissions index chart, which illustrates how admissions decisions could be made in a compensatory fashion allowing higher performance in either test scores or grades to serve a compensatory role for lower test scores or grades.
4. Plaintiffs failed to gain acceptance by automatic admission where the top 10% of students in each Texas high school's graduating class were accepted to the University of Texas at Austin, irrespective of race, and were denied admission during a second review for students not automatically admitted. In 2009, legislation limited automatic admissions to 75% of the university's enrollment capacity, which has lowered the automatic admission to the top 6% at each high school (Jaschik, 2017).
5. Of the 58 items, 38 prohibit calculator use.
6. The last official concordance study sanctioned by both testing programs and tables can be found at ACT (2018).
7. The ACT was administered exclusively on computer for international test takers, and available by computer for some state and district school-based administrations prior to 2024.
8. Open admission is an unselective and noncompetitive college admissions process in the United States, where the only requirement is a high school diploma or a General Educational Development certificate. The number of open-admission institutions was obtained through the National Center for Education Statistics Integrated Postsecondary Education Data System data center at https://nces.ed.gov/ipeds/datacenter/reportmain.aspx. Sorting variables for the 2018–2019 year included admissions and test scores, open admission policy (yes); and the following institutional characteristics: level of institution (1 four or more years); sector of institution (1 public, 4 years or above; and 2 private, not-for-profit, 4 years or above); degree-granting status (1 degree granting); and undergraduate offering (1 undergraduate degree or certificate offering). Total enrollment was about 2 million at these open-admission institutions. Some of the institutions were branch campuses of institutions that had selective admissions at their main campus.

9. Includes individuals ages 16–24 who graduated high school or completed a General Educational Development certificate or other equivalency program. See https://nces.ed.gov/programs/coe/indicator_cpa.asp.
10. The ACT and SAT report degree aspirations separately and with slightly different categories. The percentages were calculated by a weighted average from ACT and SAT data, which may result in double counts of students who completed both tests.
11. The National Collegiate Athletics Association suspended and has now eliminated its test score requirement for Division I or II athletics (Bauer-Wolf, 2023).
12. After July, 2024, the name transitioned back to the *GMAT Exam* (Graduate Management Admission Council, 2024a).
13. Two fixed-length paper forms called companion tests had been developed for each of the five adaptive tests.