

2 DEVELOPMENT AND EVOLUTION OF THE SAT AND ACT

Michelle Croft and Jonathan J. Beard¹

College admissions tests were initially developed as a mechanism to standardize the college admissions process and provide students greater access to highly selective institutions. Over time and in response to societal changes, the testing programs have shifted to accommodate a broader range of college applicants as well as a broader range of uses of the test scores. As the origins of the SAT and the ACT test (hereinafter "ACT") have been well documented, this chapter will provide a brief summary on the origins of the two assessments but primarily focus on these shifts in society and in the testing programs, as well as on access to postsecondary education and to the tests themselves.

The rest of this chapter is organized using broad themes that will structure the discussion of college admissions tests and their place in the modern era: assessment origins; assessment transparency (including test preparation, test security, truth in testing); detection and removal of cultural bias; the use of college admissions tests for K-12 federal accountability; and access for students with disabilities and English language learners. Last, we offer a brief discussion of lessons learned from the constraints and demands placed on the testing profession.

SAT Development

At the time of the inception in November 1900 of The College Entrance Examination Board, test developer of the SAT, there was effectively no uniformity in the college admissions process in terms of both the subject matter content needed and the degree of proficiency required to be successful in postsecondary education. It was difficult for students to know the standard that would be applied to them for college admissions. For college administrators, the variability among secondary schools and students led to the development of unique sets of exams in multiple

subjects for each institution. The College Board exams were designed to bring a standard of uniformity to the selection process. The content of the exam was to be uniform in terms of subject matter, the exams were to be uniformly administered in terms of date and exam timing across multiple locations, and all responses were to be scored in a uniform way to the same standard. The exam results would serve to provide some economy of force with respect to admissions decisions and clarity of expectations on the part of students.

The production of the exams was somewhat secondary to the overarching goal of providing a means of communication between schools and colleges moving toward a more uniform secondary school curriculum (Angoff & Dyer, 1971, p. 1). However, without exams serving as a unifying aspect, such communication and uniformity were unlikely to be achieved. Although there was some tension among schools, colleges, and the Board regarding curricula and course content, the implementation of the exams reflected a genuine if ambitious desire to bring "law and order into an educational anarchy" with regard to the subject matter of postsecondary preparation and the degree to which the subject matter should be mastered (Angoff & Dyer, 1971, pp. 1-2).

The first "College Boards" were given in 1901 as all-essay responses in the subjects of English, French, German, Latin, Greek, history, mathematics, chemistry, and physics, given over the course of several days. In the next iteration, Spanish, botany, geography, and drawing were added. Scores were given as ratings: Excellent, Good, Doubtful, Poor, and Very Poor. Corresponding percentage scores were given as 100-90, 89-75, 74-60, 59-40, and less than 40. While initially designed to test students' factual subject knowledge, by 1925 there had been a pivot to an approach in which students would apply general knowledge and working principles to novel situations (Angoff & Dyer, 1971, p. 2). The change prompted some pushback on the part of stakeholders out of concern that the changes would remove the emphasis from subject matter knowledge and mastery to more of a "superficial cleverness" (Angoff & Dyer, 1971, p. 2; see also, Donlon, 1984; Lawrence et al., 2003; Zwick, 2002).

The first multiple-choice Scholastic Aptitude Test (SAT) was administered in 1926, to just over 8,000 examinees. The change was, in part, to be cost- and time-efficient, as well as to remove subjective judgment in the scoring process (see e.g., Trachsel, 1992.) Scores were given on nine subtests: Definitions, Arithmetical Problems, Classification, Artificial Language, Antonyms, Number Series, Analogies, Logical Inference, and Paragraph Reading. Although the name and intent of the test was to "[distinguish] it from tests of achievement in school subjects" (Angoff & Dyer, 1971, p. 2), the developers cautioned that the test was *not* designed to measure general intelligence or "mental alertness." Even at its humble outset, discussions of the college admissions test included warnings about placing too great an emphasis on test scores, stating that to do so "is as dangerous as the failure [to] properly evaluate any score or rank in conjunction with other measures" (Angoff & Dyer, 1971, p. 2). The SAT score was best regarded as a complement to the applicant's educational experience.

Between 1929 and 1941, many of the most familiar features of the SAT were established with a few broad changes to the test:

1. In April 1929, the test was divided into two major sections: verbal and mathematical.
2. In 1937, a second administration of the SAT was offered in June.
3. In 1941, the SAT scale (200–800) was established using the April administration (about 11,000 examinees).
4. Also, in 1941, the regular procedure of equating June scores to the April form was introduced.

Other notable changes (e.g., to test parts and allotted time) occurred over the much longer period of 1941–2002. A summary of all the changes to the SAT is shown in Figure 2.1.²

Throughout the SAT's evolution, it should be noted that early writings about the SAT indicate that *something* other than *specific* content mastery was part of the test. The best description regarding the intent of the SAT to complement a student's educational experience is provided by Donlon & Angoff (1971):

The SAT was, in a sense, intended to provide some redress for possible errors and inconsistencies in secondary school records and in the old essay examinations that were tailored to specific curriculums. By stressing the direct measurement of basic abilities, the rationale was that it would offer an opportunity for a more balanced assessment of the student who had failed to achieve subject-matter mastery in keeping with his development of these basic abilities.

(p. 15)

This line of thinking is understandable, as the SAT was *part* of College Board's Admission Testing Program, consisting of the SAT and several achievement tests. The achievement tests were much more focused on content mastery within a very specific domain. Fremer and Chandler note that

the usefulness of the SAT as an indicator of a student's potential for college work depends in large measure on the fact that the SAT *measures general ability as it has developed over the full range of experiences in a person's life* (1971, p. 147, *emphasis added*)

In a similar vein, Coffman notes that the SAT was designed to identify students who possess the skills needed to do college-level work regardless of what they may have studied in high school (1971, p. 49).

At the outset, the goal of having a test composed of items that measured attainment in a given domain, without being overly contingent on the specific

FIGURE 2.1

Year (19XX)	Antonyms	6-choice	5-choice	Parts of speech	Analogies	Select 4th term	Select 2 terms	Select 2nd pair	Paragraph reading	Reading comprehension	Critical reading	Double definition	Sentence completion	Definitions	Classification	Artificial language	Logical inference	Synonyms (2 choice)	Math	Problems: fill in	Problems: 6-choice	Problems: 5-choice	Arithmetic word problem	Data sufficiency	Number series completion	Quantitative comparison	Student produced response
26	x																										
27	x																										
28	x																										
29	x																										
30	x																										
31	x																										
32	x																										
33	x																										
34	x																										
35	x																										
36	x																										
37	x																										
38	x																										
39	x																										
40	x																										
41	x																										
42	x																										
43	x																										
44	x																										
45	x																										
46	x																										
47	x																										
48	x																										
49	x																										
50	x																										
51	x																										
52	x																										
53	x																										
54	x																										
55	x																										
56	x																										
57	x																										
58	x																										
59	x																										
60	x																										
74	x																										
75	x																										
86	x																										
94	x																										
95	x																										
96	x																										

FIGURE 2.1 Summary of changes to the SAT

curriculum a student may have been exposed to, was appealing; developing item types that function well for that purpose would be considered a worthy endeavor. However, the goal of having items that functioned well psychometrically, and were not completely subject to differences in formal instruction, did not come without a cost. A natural tension seemed to exist between the traits that were being measured and the types of items that were included on previous versions of the SAT. On the one hand, the questions were written in such a unique way that a reasonable amount of content knowledge and problem-solving ability had to be brought to bear to answer them.

On the other hand, *because* of their relative uniqueness in terms of how the questions were asked, these item types lent themselves to being somewhat coachable (Gladwell, 2001; Slack & Porter, 1980a; Jackson, 1980; Slack & Porter, 1980b; DerSimonian & Laird, 1983). However, it is important to note that having students become familiar with the nature of the test was fundamental to the early SAT. From 1926 through 1944, students were *required* to show a completed practice test in order to take the exam for admission purposes (Fremer & Chandler, 1971). After 1944, examinees no longer had to demonstrate completion of a practice test, and descriptive information and sample questions were made available to students.

As noted in Figure 2.1, many changes and modifications have been made to the SAT testing program. A number of the changes have moved the test further from “puzzle-solving” question types to ones that require knowledge and reasoning skills closely related to those experienced in school coursework (Lawrence et al., 2003). With each iteration, the test changed in a way such that the interpretations of the test scores could be more informative and ultimately more useful for test score users. Some of the changes were in response to external demands or constraints, whereas other updates were based on changes in curricular emphasis. Lawrence et al. (2003) provide an exceptionally well written summary of the changes made to the test up to that time, and the reasoning for why particular modifications were made.³

The latest redesign of the SAT in 2016 encompassed a comprehensive and deliberate move toward clear, specific, and transparent aspects of the SAT testing program (College Board, 2015b). Broadly, the redesign removed the required essay component that had been added in 2005, added subscores, reduced answer options from five to four, and ended the scoring practice in which fractions of points were subtracted for wrong answers (i.e., formula scoring; College Board, 2015b). Lastly, a distinctive element of the 2016 redesign was the change in test specifications to focus more on rigorous coursework and problem solving in a college-readiness context (College Board, 2015a). Lawrence et al. (2003) succinctly articulate the importance of change when necessary: any testing program must strive to be as fair as possible to examinees and as informative as possible for college admissions decisions. The latest changes to the SAT with respect to transparency, as well as near universal access to practice and review materials, meet both goals of fairness to students as well as usefulness to colleges.

ACT Development

While use of the SAT had become common at highly selective postsecondary institutions in the northeastern United States during the 1950s, many regions and institutions were without a standard college admissions test. This situation became especially problematic after World War II, when there was an increase in the number of postsecondary applicants, and these applicants were much more diverse than those attending elite institutions using the SAT (Lazerson, 1998). The increase was due in part to the growing US economy, an increase in the population generally, and additional federal money being spent on higher education that made it possible for more students to attend college.

For instance, the G.I. Bill of 1944 (Servicemen’s Readjustment Act, Pub. L. 78–346) expanded opportunity to higher education by providing veterans with financial aid. Similarly, the National Defense Education Act (NDEA), enacted in 1958 (Pub. L. 85–864), provided low-interest loans for higher education and eventually evolved into the National Direct Student Loan Program and the Perkins Loan Program. Further, the Great Society programs of the 1960s included additional federal initiatives to expand access to higher education. For example, the Economic Opportunity Act of 1964 (Pub. L. 88–452) authorized grants that became college access programs such as Upward Bound, which provides college admissions support to high school students from low-income families and from families in which neither parent holds a bachelor’s degree. The Economic Opportunity Act also funded Work/Study Programs to promote part-time employment of college students from low-income families in need of earnings to help cover college costs (Pub. L. 88–452, Sec. 121).

In response to the increase in college applicants, states started to create their own admissions tests as the SAT was not marketed toward those institutions (ACT, 2009). The tests were of varied quality and made it difficult for out-of-state students interested in attending other states’ colleges to participate. In addition to assisting with admissions decisions, universities needed assistance with placement decisions once the student enrolled. There was also a concern that aptitude tests—as the SAT was designated at the time—were not the appropriate tool for determining college admissions on a large scale. Instead, there should be a tool that could “gauge students’ readiness and ability to perform college-level activities” (ACT, 2009, p. 9).

For these reasons, the ACT was developed.⁴ It grew out of the Iowa Academic Meet, which was the first Iowa test program for high schools. The Meet included multiple rounds of testing to identify outstanding scholars, and the first round—the “Every-Pupil Test”—was administered to every high school student enrolled in the tested subject. The purpose of assessing all students was not only to identify outstanding scholars but also to raise the standards of instruction and to interest students in the subject matter (Lindquist, 1976).

The Iowa Academic Meet eventually grew into the Iowa Tests of Basic Skills and the Iowa Test of Educational Development (ITED). The new tests were

designed to emphasize “the development of skills and generalized abilities, as opposed to rote learning of subject matter” and to require problem solving and critical thinking (ACT, 2009, pp. 2–3).

The ACT was first administered in 1959. The pace of its development was supported by the ability to leverage the ITED and pre-test items during the administration of the ITED. The organization also benefited from the technical capability of a newly developed optical scanner, which had the ability to process tens of thousands of answer sheets a year. The first ACT included four sections: English, Mathematics, Social Studies, and Science. The Social Studies and Science tests emphasized reasoning and problem solving through the interpretation of readings in the relevant subjects. The test took three hours, with 45 minutes for each section, and the scores were reported on a 0-to-36 scale, with a mean composite score for college admissions purposes.

The ACT grew quickly. In the 1959–60 school year, 132,963 students took the test (ACT, 2009); in 1962–63, 368,943 students took it; and 961,184 students took the test in 1967–68.

Over the years, there were minor changes to the ACT. After reviews of textbooks for grades 7–12 and interviews with experts, there were multiple, more significant, changes to the test in 1989 (ACT, 2009): it was updated to provide subscores that could be helpful for course placement purposes; the Social Studies test was replaced by a Reading test that measured “pure” reading ability (ACT, 2009, p. 64); and the Science section was changed to focus on the scientific process. ACT continues to update the test based on changes in high school curriculum and what is needed for college. Currently, the ACT uses a National Curriculum Survey administered approximately every three to five years to identify the most current college-level coursework expectations and high school curriculum content (ACT, 2016).

Transparency: Test Preparation, Test Security, and Truth in Testing

Despite the growing number of students taking the SAT and ACT post–World War II, students—and the general public—did not know much about the tests. Information about scoring procedures and validity studies was not generally released (Robertson, 1980). Students were not privy to copies of their answers after the tests, nor were the tests themselves publicly released. The lack of information about the tests led to the rise of test-preparation companies, many of which were making false and misleading advertising claims (see Haney, Madaus, & Lyons, 1993; Zwirk, 2002). Ultimately, the Federal Trade Commission conducted an inquiry related to SAT and Law School Admission Test preparation companies in 1978.

One outcome of the rise of test-preparation companies was the implementation of test security measures. Although some test prep companies published collections of items that were only purportedly included on the tests, others

published items taken directly from the tests, raising concerns about test security (Angoff & Dyer, 1971). As a result, College Board began to implement procedures to secure the SAT item pool (Angoff & Dyer, 1971). Some of these measures included the use of sealed, numbered item booklets and of procedures around shipping and storing the booklets. In instances where scores were questioned, students were offered the chance to retake the test to confirm the scores received originally. ACT introduced similar procedures at about the same time (ACT 2009, p. 64). Since the introduction of test security measures, both organizations have continued to refine and adapt test administration policies with the goal of preventing test irregularities; however, as not all irregularities can be prevented, both organizations have adopted additional mechanisms to detect breaches if they do occur.

Another outcome was the introduction of truth-in-testing laws. Two states—New York and California—enacted legislation which required test providers to provide more information about the tests to the public.⁵ New York’s law was passed in 1979 and required providers to file with the New York Commissioner of Education “all test questions used in calculating the test subjects’ raw score” and “the corresponding acceptable answer to those questions,” as well as data about test takers (NY Educ. Code § 341, 342). College Board, along with the Graduate Management Admissions Council, the Test of English as a Foreign Language Policy Council, and the Educational Testing Service, filed a lawsuit in 1990 for a preliminary injunction claiming that the law violated the Copyright Act of 1976 (*College Entrance Examination Bd. v. Pataki*, 889 F.Supp.554, 1995 ND.N.Y.).⁶

The plaintiffs claimed that the disclosure provisions had forced them to reduce the yearly number of test dates offered in New York—even as they increased the number of dates offered in all other states—because the reduction would require them to release a smaller number of test forms each year. On a motion for reconsideration, the court specified the number of forms that must be released annually (*College Entrance Examination Bd. v. Pataki*, 893 F.Supp. 152, NYND, July 26, 1995). The court decision was codified in statute in 1996 (NY Educ. Code § 342). The statute was later changed to require the release of either two-thirds of tests administered in the test year or, for the entities that were included in the lawsuit, a specific number of test forms depending on the number of test dates offered in the state.

California enacted a truth-in-testing law in 1978 (Cal Ed. Code § 99150–99160). The original version of the law differed from New York’s in that test providers could file with the state educational commission either a completed sample test or list of “representative” test questions and answers as opposed to actual test forms (Cal. Educ. Code § 99152; Robertson, 1980). Later, the law was changed to require release of a certain proportion of test forms (Cal. Educ. Code § 99157).

In addition to release of test content (either complete or representative), both states required disclosures related to scoring methodology and reporting

(Robertson, 1980; The Educational Testing Act of 1979, 1980). California also required financial disclosures about fees and expenses for each testing year.

Although ACT was not a party to the 1990 lawsuit, as a result, it reduced the number of its test dates in New York to comply with the law (ACT, 2009, p. 50). It was not until 2019 that a minor update to the law's release requirement allowed ACT to offer an additional test date (New York S.B. 8639, 2018; ACT, 2018).

Additionally, in response to these laws, College Board and ACT both began releasing items and test forms not just to students in California and New York but nationally. College Board made a sample set of questions available to the public in 1978 (Valley, 1992), and a year after allowing New York SAT examinees to receive a copy of their exam, answer sheet, and correct answers in 1980, it extended the opportunity to examinees in other states (Fiske, 1981). Similarly, in December 1982, ACT began selecting national test dates when students could receive copies of their test questions and answers (ACT, 2009).

After first publishing a booklet that included sample questions and general test-taking information in the early 1980s, ACT began to share items with various publishers for the creation of test preparation materials (ACT, 2009). In 1985, ACT offered a free publication, *Preparing for the ACT Assessment*, which included a full test form with answer key and a worksheet to help students analyze their strengths and weaknesses on each section of the test.

Both organizations have continued to offer free test prep materials and have expanded free test prep opportunities. For example, College Board partners with Khan Academy to offer personalized study plans based on students' performance on past SAT results or diagnostic quizzes (College Board, n.d.), while ACT offers a free online learning tool and test practice program (ACT, n.d.).

The rise of test preparation materials issued by College Board and ACT has also raised criticism. Some critics argued that previous versions of the SAT were not truly an aptitude test, because test prep companies had been coaching students on particular SAT items. College Board and other researchers conducted seven studies investigating short-term coaching and found only small and insignificant gains (Angoff & Dyer, 1971; Messick et al., 1980).⁷ ACT's position was that, because the ACT tested students' familiarity and facility with what they learned in high school, coaching could not provide the content knowledge necessary to make a difference in a student's score. However, "familiarity with the test format, procedures, and test-taking skills," along with reviewing content knowledge that had been learned several years before taking the test, did have an impact on a student's score (ACT, 2009, p. 58). In any case, it is reasonable to conclude that, before truth-in-testing laws, test prep that exposed students to the type of content tested could have impacted scores, not because the students had greater knowledge or abilities, but simply because they had been exposed to the type of content to be tested.

Detection and Removal of Cultural Bias

The lack of transparency around what was tested may have also led to perceptions that the tests were culturally biased. Some of the criticism may have been due to lack of information about the tests, particularly as reports regarding validity were not widely available. Other criticism may have been in part due to score gaps noted between majority and minority students (Fallows, 1980; Weber, 1974; The Educational Testing Act of 1979, 1980).

In order to address these concerns, throughout the 1970s and 1980s, both College Board and ACT refined their policies for reviewing items for cultural bias. By 1980, College Board had replaced its informal policies related to detecting item bias with a formal policy that included specific guidelines for test development (see Valley, 1992; Dorans, 2013; and Zwick, 2002). The guidelines provided specific instructions to item writers to include material reflecting different cultural backgrounds as well as to avoid material that may be potentially offensive.

In addition, several early approaches to quantifying item bias, such as Angoff's delta plot method, were introduced for the SAT (see Angoff, 1972); by the late 1980s, and in response to a settlement in the lawsuit *Golden Rule Life Insurance Company v. Illinois Insurance Director and Education Testing Service*, these approaches had been replaced with differential item functioning (DIF) analysis (see Faggen, 1987). Broadly, DIF analysis matches examinees of different subgroups (e.g., race/ethnicity or gender) on their level of knowledge and skills and then compares the performance of different groups on test items (Dorans & Kulick, 1986; Holland & Thayer, 1988; Dorans & Holland, 1992; and Dorans, 2013). If DIF is not present, the likelihood of answering a particular item is dictated by the level of ability a student has, not any other characteristic (e.g., race/ethnicity). College Board and ACT also conducted validity work to examine the predictive validity evidence for different racial/ethnic groups (see, e.g., Cleary, 1968; Maxey & Sawyer, 1981; and Dorans, 2013).

ACT, which had conducted content and bias review panels since its inception, "engaging representatives from various minority groups to review the language and content of questions" (ACT, 2009, p. 58), added a second process in 1981 to detect potential bias using statistical reviews of data from operational items. This process, like College Board's early approaches, was eventually replaced by DIF analysis.

Use Within High Schools: Standards, Relevance, and Accountability

As the tests were redesigned to be more reflective of the high school curriculum, more states were interested in using the tests as part of their accountability systems. The shift started during the 2001 No Child Left Behind (NCLB)

reauthorization of the Elementary and Secondary Education Act (ESEA) (Public Law 107–110) but increased after the introduction of the ESEA Waiver program in 2011 and the subsequent reauthorization of ESEA, the Every Student Succeeds Act (ESSA) in 2015.

NCLB required states to adopt content standards and aligned assessments in grades 3–8 and once in high school for reading and mathematics, as well as science for certain grade spans (20 U.S.C. § 1111(b)(3)). NCLB gave states discretion in terms of the rigor of the state's content standards, but the general emphasis was on proficiency (Hoff, 2002). During this time, three states—Illinois, Maine, and Michigan—opted to use the SAT or ACT as part of the state's high school assessment system (Camara et al., 2019).

With NCLB's emphasis on proficiency levels, there was growing dissatisfaction with its implementation. By 2011, four out of five schools were not expected to meet proficiency goals within the next year, raising concerns that many schools could not meet the law's goal of 100 percent proficient by 2014 (Duncan, 2011; Obama, 2011). Although ESEA was due for Congressional reauthorization, there was not much movement to do so. As a stopgap, the Obama administration created the ESEA Waivers, under which qualifying states would have greater flexibility in designing their own accountability systems to enable greater student proficiency in ELA, math, and science (USED, 2016a). One of the terms states needed to meet to receive a waiver was adoption of more rigorous college- and career-ready academic standards.

The change to college and career readiness expectations was partially a recognition that state standards adopted under NCLB were not sufficiently rigorous (National Center for Education Statistics, 2007; Byrd Carmichael, Wilson, Porter-Magee, & Martino, 2010). It was also an acknowledgment of an effort led by the National Governors Association and the Council of Chief State School Officers to develop common, research- and evidence-based content standards that would help prepare students for college and careers (Common Core State Standards Initiative, n.d.). The Common Core State Standards working group included experts from both College Board and ACT to provide information on postsecondary readiness (National Governors Association & The Council of Chief State School Officers, 2009). At its peak, all but four states adopted the Common Core State Standards (Ujifusa, 2015a), and although a number of states would subsequently change their standards, the changes have largely been minor (Friedberg et al., 2018). Because of the ESEA Waivers and state adoption of the Common Core State Standards, the majority of states would eventually have similar standards that emphasized college and career readiness.

In addition to changing academic standards, there was also mounting political pressure to reduce student testing time. After the Common Core State Standards were developed, the Obama administration funded the development of tests aligned to the standards (US Department of Education, 2010). Two state-led consortia emerged in response to the funding opportunity, and each eventually

developed its own test: The Partnership for Assessment of Readiness for College and Careers (PARCC) assessment, and the Smarter Balanced Assessment Consortium (Smarter Balanced) assessment. Both of the tests had significantly longer administration times than previous state tests (Doorey & Pollikoff, 2016, p. 29; Camara et al., 2019). In high school in particular, test length was an issue because students were also preparing for other tests, such as college admissions tests and Advanced Placement tests, that more directly affected their lives after high school. After the introduction of the two tests, there was a rise in the number of parents opting their children out of taking them, particularly at the high school level (see, e.g., Bennett, 2016; Croft, 2015; Golden & Webster, 2019). A number of states received letters from the US Department of Education about their low student participation rate in statewide testing, which could impact their ability to continue receiving federal funding (Ujifusa, 2015b). Therefore, by the time ESSA was enacted, states were looking for ways to decrease testing at all grade levels, but particularly in high school.

In addition to issues related to student participation in testing, there were also changes to accountability systems more generally. For example, where NCLB had been rigid in terms of criteria for identifying schools in need of improvement and the consequences and supports applied to schools once they were so identified, ESSA gave states leeway on in determining such criteria (Lyons, D'Brot, & Landl, 2017). Also, while student academic achievement remained a significant component of a state's required accountability system, ESSA added other components, including a school quality and/or student success indicator. Further, the law explicitly allowed districts—with state approval—to administer a locally selected, nationally-recognized test such as the SAT or ACT in place of the state's chosen high school academic achievement assessment (20 U.S.C. § 1111(b)(2)(h)).

In some cases, the state's chosen assessment was the SAT or ACT. Since 2001, a number of states had already been administering one of the tests to all of their eleventh graders during the school day and at no cost to the student, as a way to encourage more of them to apply to college. Under ESSA, some states chose to carry this practice forward as their academic achievement indicator to increase the quantity and quality of test participation (Camara et al., 2019; Marion, 2018) as well as to potentially improve college-going rates (Hurwitz et al., 2015; Hymn, 2017). In the 2018–2019 school year, a total of 30 states administered the SAT or ACT in eleventh grade as either a state requirement or as an optional test for which the state reimburses individual districts that choose to administer it (Croft, Vitale, & Guffy, 2019). Thirteen states used the SAT or ACT as the state's academic achievement indicator under ESSA, and 26 states used the SAT and/or ACT under ESSA as an acceptable measure of postsecondary readiness within the school quality/student success indicator.

The use of SAT and ACT scores for accountability has not been without controversy. Critics contend that the tests were not designed with accountability in mind and that they are not sufficiently aligned to state content standards to be

appropriate for such use (Marion & Domaleski, 2019; Achieve, 2018). However, others contend that there is sufficient validity evidence to support it (Camara et al., 2019).

As of this writing, the SAT and ACT have each substantially met the US Department of Education's peer review requirements in at least one state (Brogan, 2018; Brogan, 2019). However, additional alignment evidence will still be needed before each test can fully meet requirements.

Access for Students with Disabilities and English Learners

Given the expansion of testing to all students within a state, it is increasingly important that the tests are accessible and accurately reflect what students know and can do. Federal law has historically pushed testing companies toward expanding access for students with disabilities by providing accommodations. Prior to federal involvement, College Board and ACT were already providing accommodations such as additional time, use of a typewriter, and Braille versions of the tests for students with physical disabilities (Donlon & Angoff, 1971). However, testing organizations cautioned those who used the scores to make admissions decisions not to place substantial importance on them, because not enough research had been done to support their comparability with the scores of examinees who had taken the test without the accommodations. Instead, the test users were encouraged to weigh accommodated students' previous academic records more heavily in their decision making than those of non-accommodated students (Donlon & Angoff, 1971; Laing & Farmer, 1984). As the research base grew, this caution—also known as flagging—was given only when the accommodation included extended time (Lewin, 2000, 2002).

Accommodations were not required by federal law until enactment of the Rehabilitation Act of 1973 (20 U.S.C. § 701 et seq.), which prohibited the use of tests that have an adverse impact on persons with disabilities, including learning disabilities⁸ (34 C.F.R. 104.42(b)(2)). The Act required that

when a test is administered to an applicant who has a handicap that impairs sensory, manual, or speaking skills, the test results accurately reflect the applicant's aptitude or achievement level or whatever other factor the test purports to measure, rather than reflecting the applicant's impaired sensory, manual, or speaking skills (except where those skills are factors that the test purports to measure)

(34 CFR 104.42(b)(3)(i))

The Rehabilitation Act did not directly require College Board and ACT to provide appropriate accommodations themselves but, rather, regulations enacted in 1980 required postsecondary institutions receiving federal funding to ensure that the admissions tests they used provided them (34 CFR 104.3). In 1990, the

Americans with Disabilities Act (ADA) (42 U.S.C. § 12189; United States Department of Justice, 2014) further expanded opportunities for individuals with disabilities to access postsecondary education by continuing the accommodations requirements from the Rehabilitation Act.

After passage of the ADA, there was growing concern among disability rights advocates about the practice of flagging test scores for nonstandard administrations. Critics contended that the practice “raises issues of stigma, privacy, and discrimination against disabled examinees” (Mayer, 1998). The criticism was not limited to the SAT and ACT. Eventually a lawsuit was filed against Educational Testing Service (ETS), the makers of the Graduate Management Admission Test (GMAT), when an examinee's test scores were flagged as a nonstandard administration because of the examinee's use of extended time and a trackball (*Breimhorst v. Educational Testing Service*, 2000). After a motion to dismiss was denied, ETS opted to settle the lawsuit and discontinue flagging GMAT scores as well as scores on the Graduate Records Exam and the Test of English as a Foreign Language (Sireci, 2005).

Although College Board was not involved in the GMAT lawsuit, it opted to convene a “Blue Ribbon Panel on Flagging” jointly with the *Breimhorst* plaintiff's attorneys in the spring of 2002 (Gregg, Mather, Shaywitz, & Sireci, 2002; Sireci, 2005). The panel consisted of six members, two of whom were psychometricians, and a nonvoting chair. The panel reviewed more than a dozen studies but could not reach consensus. In the end, a four-member majority recommended that flagging be discontinued, and College Board accepted the recommendation (Sireci, 2005). Shortly thereafter, upon reviewing its own research, ACT announced that, in fall 2003, it would no longer flag scores of students who tested with extended time (Lewin, 2002).

The laws and policies discussed here pertain only to students with disabilities. But as more states began to use the SAT and ACT as part of their accountability systems, an increasing number of English learners were being required to take the tests. There was a concern that the scores of these examinees did not accurately reflect the students' skills and knowledge but only their degree of English proficiency. However, because these students were not covered by the ADA, they were not offered accommodations. The matter of extending accommodations to English learners was complicated by the fact that being a “learner” is a temporary state and the examinee would eventually become proficient in English.

As part of ESSA negotiated rulemaking, language was proposed in April 2016 related to accommodations both for English learners and students with disabilities (Title I, Part A negotiated rulemaking, U.S. Department of Education, 2016, April 19). The “equal benefits” language required states to ensure that appropriate accommodations were provided to students with disabilities and English learners such that neither group would be denied the benefits of participation in the test; namely, a college reportable score (34 C.F.R. 200.3(b)(2)(i)). The regulations were published for public comment in July 2016 and were finalized in December

2016). Like the Rehabilitation Act, the regulations are not directly applicable to College Board and ACT; rather, in this case, they apply to any state seeking to use the scores as part of its accountability system.

Soon after the regulations were released for public comment, ACT convened “a panel of external experts representing state education agencies, colleges, English learner and bilingual policy administrators from state departments of education, civil rights advocates, testing and measurement experts, and researchers” to help determine whether and what supports could be offered to English learners without violating the tested constructs and thus invalidating the college report-ability of their scores (ACT, 2016; Moore, Huang, Huh, Li, & Camara, 2018). The panel recommended providing four testing supports for English learners: additional time; use of an approved word-to-word bilingual glossary; test instructions in the student’s native language; and testing in a non-distracting environment. Shortly after ACT announced the availability of English-learner supports, College Board also did so (College Board, 2016).

Discussion

Societal demands and shifts in policy helped establish the role of the SAT and ACT and have ultimately contributed to their ongoing evolution and improvement. The programs have expanded their scope, and the tests are used by a broader and more diverse range of students and postsecondary institutions than before. At the same time, the test content has become more aligned with high school curricula and the content needed for college. The testing companies also established mechanisms to check for bias and made changes to provide more transparency into the testing process through the release of test forms and broader sharing of general information about scoring and predictive validity. As education policy continues to evolve, we expect that there will continue to be shifts and refinements in the programs.

Notes

- 1 The authors would like to thank Michael Walker, Wayne Camara, and Dan Vitale for their thoughtful reviews of an earlier draft of this chapter.
- 2 More detailed information regarding changes and modifications can be found in Angoff, 1971; Valley, 1992; Dorans, 2002; and Lawrence et al., 2003—although the reader should note that there are some minor inconsistencies across the three, especially between Angoff and Lawrence et al. The latter also contains an excellent summary of the included items and testing time allotted (Lawrence et al., 2003, pp. 6, 10).
- 3 For example, the 1994 revision introduced calculators on certain multiple-choice items and on grid-in item types for the math section. The 1994 revision also removed antonyms. In addition to changes in item type, the Math and Verbal scales were recentered in 1995 (Dorans, 2002). Similarly, in 2005 changes were made to add an essay and remove analogies. Readers are encouraged to read the original Lawrence et al. (2003) report for detailed information.

- 4 For a detailed description of the history of ACT, please see ACT, 2009.
- 5 In addition, Colorado, Florida, Hawaii, Maryland, Ohio, Pennsylvania, and Texas proposed truth-in-testing bills, and there were two proposed federal truth-in-testing bills (Robertson, 1980).
- 6 The lawsuit was based on a similar lawsuit by the American Association of Medical Colleges (928 F.2d 519, 2nd Cir.), which was filed in 1979 and took nearly ten years of litigation before the court ruled for the plaintiff.
- 7 See Messick et al. for an excellent summary of additional studies.
- 8 In the 1985–1986 school year, 75 percent of the 5,000 students who received accommodations on the ACT had been classified as learning disabled (ACT, 2009, p. 60).

References

- Achieve (2018, March). *What Gets Tested Gets Taught: Cautions for using college admissions tests in state accountability systems*. Achieve. <https://www.achieve.org/files/CollegeAdmissionsExamBrief2018.pdf>.
- ACT (2009). *ACT: The First Fifty Years, 1959–2009*. ACT, Inc.
- ACT (n.d.). *Free ACT Test Prep*. <https://www.act.org/content/act/en/products-and-services/the-act/test-preparation/free-act-test-prep.html>.
- ACT (2016). *ACT National Curriculum Survey 2016*. ACT, Inc. http://www.act.org/content/dam/act/unsecured/documents/NCS_Report_Web.pdf.
- ACT. (2018, Nov. 6). ACT to add February ACT test date in New York state. *ACT Newsroom & Blog*. <http://leadershipblog.act.org/2018/11/act-to-add-february-act-test-date-in.html>.
- The Americans with Disabilities Act of 1990, Pub. Law 101–336, 42 U.S.C. § 12101 et seq.
- Angoff, W. H. (1972, September). *A technique for the investigation of cultural differences*. Paper presented at the meeting of the American Psychological Association, Honolulu, HI. <https://files.eric.ed.gov/fulltext/ED069686.pdf>.
- Angoff, W. H., & Dyer, H. S. (1971). The Admissions Testing Program. In W. H. Angoff (Ed.), *The College Board Admissions Testing Program: A technical report on research and development activities relating to the Scholastic Aptitude Test and Achievement Tests* (pp. 1–13). College Entrance Examination Board.
- Bennett, R. E. (2016, April). Opt out: An examination of issues. *ETS Research Report Series*. <https://doi.org/10.1002/ets2.12101>.
- Breimhorst v. Educational Testing Service, C-99-3387 WHO (N.D. Cal. 2000)
- Brogan, F. (2018). *Letter to Superintendent Evers*. <https://www2.ed.gov/admins/lead/account/nclbfinalassess/vt7.pdf>.
- Brogan, F. (2019). *Letter to Superintendent Rice*. <https://www2.ed.gov/admins/lead/account/nclbfinalassess/michigan9.pdf>.
- Byrd Carmichael, S., Wilson, S. W., Porter-Magee, K., & Martino, G. (2010, July 21). *The State of State Standards—and the Common Core—in 2010*. Thomas B. Fordham Institute. <https://fordhaminstitute.org/national/research/state-state-standards-and-common-core-2010>.
- Cal Ed. Code § 99150–99160.
- Camara, W., Mattem, K., Croft, M., Vispoel, S., & Nichols, P. (2019). Validity argument in support of the use of college admissions test scores for federal accountability. *Educational Measurement: Issues and Practice*, 38(4), 12–26.
- Cleary, T. A. (1968). Test bias: Prediction of grades of Negro and White students in integrated colleges. *Journal of Educational Measurement*, 5(2), 115–124.

- Coffman, W. E. (1971). The Achievement Tests. In W. H. Angoff (Ed.), *The College Board Admissions Testing Program: A technical report on research and development activities relating to the Scholastic Aptitude Test and Achievement Tests* (pp. 49–77). College Entrance Examination Board, College Board (2015a). *Compare SAT Specifications*. College Board. <https://collegereadiness.collegeboard.org/sat/inside-the-test/compare-old-new-specifications>.
- College Board (2015b). *Test Specifications for the Redesigned SAT*. College Board <https://collegereadiness.collegeboard.org/pdf/test-specifications-redesigned-sat-1.pdf>.
- College Board (2016, Dec. 5). *Press release: College Board simplifies request process for test accommodations*. <https://www.collegeboard.org/node/24406>.
- College Board (n.d.) SAT practice on Khan Academy. <https://collegereadiness.collegeboard.org/sat/practice/khan-academy>.
- College Entrance Examination Bd. v. Pataki, 893 F.Supp. 152 (N.Y.N.D. 1995)
- College Entrance Examination Bd. v. Pataki, 889 F.Supp. 554 (N.D.N.Y. 1995)
- Common Core State Standards Initiative (n.d.) Development process. <http://www.corestandards.org/about-the-standards/development-process/>.
- Croft, M. (2015). *Opt-outs: What is lost when students do not test*. ACT. https://www.act.org/content/dam/act/unsecured/documents/5087_Issue_Brief_Opt_Outs_Web_Secured.pdf.
- Croft, M., Vitale, D., & Guffy, G. (2019, April). *Models of using college entrance examinations for accountability*. Paper presented at the National Council on Measurement in Education Annual Meeting, Toronto, ON.
- DeSimionian, R., & Laird, N. M. (1983). Evaluating the effect of coaching on SAT scores: A meta-analysis. *Harvard Educational Review*, 53(1), 1–15.
- Donlon, T. F. (1984). *The College Board Handbook for the Scholastic Aptitude Test and Achievement Tests*. College Entrance Examination Board Publishing.
- Donlon, T. F., & Angoff, W. H. (1971). The Scholastic Aptitude Test. In W. H. Angoff (Ed.), *The College Board Admissions Testing Program: A technical report on research and development activities relating to the Scholastic Aptitude Test and Achievement Tests* (pp. 15–47). College Entrance Examination Board.
- Doorey, N., & Polikoff, M. (2016, February). *Evaluating the Content and Equality of Next Generation Assessments*. Thomas B. Fordham Institute, <https://files.eric.ed.gov/fulltext/ED565742.pdf>.
- Dorans, N. J. (2002). Recentring and realigning the SAT score distributions: How and why. *Journal of Educational Measurement*, 39(1), 59–84.
- Dorans, N. J. (2013). ETS contributions to the quantitative assessment of item, test, and score fairness. *ETS R&D Scientific and Policy Contributions Series* (ETS SPC-13–04). <http://www.ets.org/Media/Research/pdf/RR-13-27.pdf>.
- Dorans, N. J., & Holland, P. W. (1992). *DIF Detection and Description: Mantel-Haenszel and standardization*. (ETS Research Report RR-92–10). Educational Testing Service.
- Dorans, N. J., & Kulick, E. (1986). Demonstrating the utility of the standardized approach to assessing unexpected differential item performance on the Scholastic Aptitude Test. *Journal of Educational Measurement*, 23(4), 355–368.
- Duncan, A. (2011, March 9). *Winning the Future With Education: Responsibility, reform and results*. Testimony before the House Committee on Education and the Workforce. <https://www.ed.gov/news/speeches/winning-future-education-responsibility-reform-and-results>.
- The Educational Testing Act of 1979 (1980): Hearings before the Subcommittee on Elementary, Secondary, and Vocational Education of the Committee on Education and Labor, House of Representatives, 96th Congress.
- The Every Student Succeeds Act, 20 U.S.C. § 1111(b)(3).
- Faggen, J. (1987). Golden Rule revisited: Introduction. *Educational Measurement; Issues and Practice*, 6(2), 5–8.
- Fallows, J. (1980). The Tests and the “brightest”: How fair are The College Boards? *The Atlantic Monthly*, 245(2), 37–48.
- Fiske, E. B. (1981, March 27). “Truth in testing” to be nationwide. *New York Times*, Section A, p. 1. <https://www.nytimes.com/1981/03/27/nyregion/truth-in-testing-to-be-nationwide.html>.
- Fremer, J., and Chandler, M. O. (1971). Special Studies. In W. H. Angoff (Ed.), *The College Board Admissions Testing Program: A technical report on research and development activities relating to the Scholastic Aptitude Test and Achievement Tests* (pp. 147–178). College Entrance Examination Board.
- Friedberg, S., Barone, D., Belding, J., Chen, A., Dixon, L., Fennell, F. S., Fisher, D., Frey, N., Howe, R., & Shanahan, T. (2018). *The State of State Standards Post-Common Core*. Thomas B. Fordham Institute. <https://fordhaminstitute.org/national/research/state-standards-post-common-core>.
- Gladwell, M. (2001, Dec. 10). Examined life. *The New Yorker*. <https://www.newyorker.com/magazine/2001/12/17/examined-life>.
- Golden E., & Webster, M. (2019, Sept. 14). More Minnesota students opting out of state tests. *Minnesota Star Tribune*. <http://www.startribune.com/more-minnesota-students-opting-out-of-state-tests/560350012/>.
- Greggs, N., Mather, N., Shaywitz, S., & Sireci, S. (2002). *The Flagging Test Scores of Individuals With Disabilities Who Are Granted the Accommodation of Extended Time: A report of the majority opinion of the Blue Ribbon Panel on Flagging*. <https://dralegal.org/wp-content/uploads/2012/09/majorityreport.pdf>.
- Haney, W. M., Madaus, G. F., & Lyons, R. (1993). *The Fractured Marketplace for Standardized Testing*. Springer Netherlands.
- Hoff, D. (2002, Oct. 9). States revise the meaning of “proficient”, *Education Week*. <https://www.edweek.org/ew/articles/2002/10/09/06tests.h22.html>.
- Holland, P. W., & Thayer, D. T. (1988). Differential item performance and the Mantel-Haenszel procedure. In H. Wainer & H. Braun (Eds.), *Test Validity* (pp. 129–145). Erlbaum.
- Hurwitz, M., Smith, J., Niu, S., & Howell, J. (2015). The Maine question: How is 4-year college enrollment affected by mandatory college entrance exams? *Educational Evaluation and Policy Analysis*, 37(1), 138–159. <http://journals.sagepub.com/doi/full/10.3102/0162373714521866>.
- Hyman, J. (2017). ACT for all: The effect of mandatory college entrance exams on post-secondary attainment and choice. *Education Finance and Policy*, 12, 281–311. http://www.mitpressjournals.org/doi/full/10.1162/EDFP_a_00206.
- Jackson, R. (1980). The Scholastic Aptitude Test: A response to Slack and Porter’s “critical appraisal.” *Harvard Educational Review*, 50(3), 382–391.
- Lang, J., & Farmer, M. (1984). Use of the ACT assessment by examinees with disabilities. *ACT Research Report Series* no. 84. ACT. https://www.act.org/content/dam/act/unsecured/documents/ACT_RR84.pdf.
- Lawrence, I. M., Rigol, G. W., Van Esen, T., & Jackson, C. A. (2003). *A Historical Perspective on the Content of the SAT* (ETS RR-03–10). College Entrance Exam Board.
- Lazerson, M. (1998). The disappointments of success: Higher education after World War II. *The Annals of the American Academy of Political and Social Science*, 559, 64–76. www.jstor.org/stable/1049607.

- Lewin, T. (2000, April 11). Disabled student is suing over test-score labeling. *The New York Times* Section 1, page 18.
- Lewin, T. (2002, July 28). ACT ends flags on test scores of the disabled. *The New York Times*, Section 1, p. 16. <https://www.nytimes.com/2002/07/28/us/act-ends-flags-on-test-scores-of-the-disabled.html>.
- Lindquist E. F. (1976, Oct. 5). *Interview by James Beilman with Dr. E. F. Lindquist*. University of Iowa Oral History Project.
- Lyons, S., D'Brot, J., & Landl, E. (2017). *State Systems of Identification and Support Under ESSA: A focus on designing and revising systems of school identification*. Washington, DC: Council of Chief State School Officers. https://ccsso.org/sites/default/files/2017-12/State%20Systems%20of%20ID%20and%20Support%20-%20Designing%20and%20Revising%20Systems_0.pdf.
- Marion, S. (2018). *What's Wrong With High School Testing and What Can We Do About It?* National Center for the Improvement of Educational Assessment. <https://files.eric.ed.gov/fulltext/ED587495.pdf>.
- Manion, S., & Domaleski, C. (2019). An argument in search of evidence: A response to Camara et al. *Educational Measurement: Issues and Practice*, 38(4), 27–28.
- Maxey, J., & Sawyer, R. (1981). Predictive validity of the ACT assessment for African-American/Black, Mexican-American/Chicano, and Caucasian-American/White students. ACT Research Bulletin, ACT Archive.
- Mayer, K. S. (1998). Flagging nonstandard test scores in admissions to institutions of higher education. *Stanford Law Review*, 50, 469–522.
- Messick, S., Alderman, D.L., Angoff, W., Jungeblut, A., Powers, D. E., Rock, D., Rubin, D. B., Stroud, T. W. F. (1980). The effectiveness of coaching for the SAT: Review and Reanalysis of Research from the Fifties to the FTC. (ETS RR-80-08). <https://www.ets.org/Media/Research/pdf/RR-80-08.pdf>.
- Moore, J., Huang, C., Huh, N., Tianli, L., & Camara, W. (2018). Testing supports for English learners: A literature review and preliminary ACT research findings (ACT Working Paper 2018–2011). ACT, Inc. <https://eric.ed.gov/?id=ED593176>.
- National Center for Education Statistics (2007). *Mapping 2005 State Proficiency Standards Onto the NAEP Scales* (NCES 2007–2482). US Department of Education. National Center for Education Statistics.
- National Defense Education Act (NDEA), Pub. L. 85–864 (1958).
- National Governors Association and The Council of Chief State School Officers (2009). *Common Core State Standards Initiative K–12 Standards Development Teams*. http://www.corestandards.org/assets/CCSS_K-12_dev-team.pdf.
- An act to amend the education law, in relation to standardized test administration, New York S.B. 8639 (2018).
- New York Educ. Code § 341, 342.
- The No Child Left Behind (NCLB) reauthorization of the Elementary and Secondary Education Act (ESEA), Pub. L. 107–110 (2001).
- Obama, B. (2011, March 14). Remarks by the President on education. Kenmore Middle School, Arlington, Virginia. Retrieved on January 30, 2020. <https://obamawhitehouse.archives.gov/the-press-office/2011/03/14/remarks-president-education-arlington-virginia>.
- The Rehabilitation Act of 1973, Pub. Law 93–112, 20 U.S.C. § 701 et seq.
- Robertson, D. F. (1980). Examining the examiners: The trend toward truth in testing. *Journal of Law & Education*, 9, 183–193.
- Servicemen's Readjustment Act, Pub. L. 78–346.
- Sireci, S. G. (2005). Unlabeling the disabled: A perspective on flagging scores from accommodated test administrations. *Educational Researcher*, 34(1), 3–12. <https://doi.org/10.3102/0013189X034001003>.
- Slack, W. V., & Porter, D. (1980a). The Scholastic Aptitude Test: A critical appraisal. *Harvard Educational Review*, 50(2), 154–175.
- Slack, W. V., & Porter, D. (1980b). Training, validity, and the issue of aptitude: A reply to Jackson. *Harvard Educational Review*, 50(3), 392–401.
- Trachsel, M. (1992). *Institutionalizing Literacy: The historical role of college entrance examinations in English*. Southern Illinois University Press.
- Ujjfusa, A. (2015a, June 29). Map: Tracking the Common Core State Standards. *Education Week*. <https://www.edweek.org/ew/section/multimedia/map-states-academic-standards-common-core-or.html>.
- Ujjfusa, A. (2015b, Dec. 23). Education department asks 13 states to address low test-participation rates. *Education Week*. http://blogs.edweek.org/edweek/campaign-k-12/2015/12/twelve_states_asked_to_address.html.
- US Department of Education (2010). Overview information: Race to the top fund assessment program; Notice inviting applications for new awards for fiscal year (FY) 2010. Federal Register, 72 no. 68. April 9, 2010. <https://www.govinfo.gov/content/pkg/FR-2010-04-09/pdf/2010-8176.pdf>.
- US Department of Education (2016b, April 19) Title I, Part A negotiated rulemaking. <https://www2.ed.gov/policy/elsec/leg/essa/title1a-assessment-consensus-regulatory-lang.pdf>.
- US Department of Education (2016a, May 12). ESEA flexibility. <https://www2.ed.gov/policy/elsec/guid/esea-flexibility/index.html>.
- United States Department of Justice (2014). ADA requirements: Testing accommodations. Technical assistance. https://www.ada.gov/regs2014/testing_accommodations.pdf.
- Valley, J. R. (1992). *The SAT: Four major modifications of the 1970–85 era*. College Board.
- Weber, G. (1974). Uses and abuses of standardized testing in schools (Occasional Papers, No. 22). Council for Basic Education.
- Zwick, R. (2002). *Fair Game? The use of standardized admission in higher education*. RoutledgeFalmer.