

CHAPTER VIII

CONCLUSION: WORLD WAR II, INCORPORATION, AND THE FINAL OSSIFICATION

Introduction

The years of World War II occupy a position of importance in the history of the development of testing in the United States that rival those of World War I. Although no single event compares in visibility with introduction of the Alpha, events between 1941 and 1945 played a crucial role in shaping the direction of testing in the postwar years. During the global conflict and in the years immediately following the armistice, the actors, script and stage that had precluded the formation of a single national testing agency in 1937 changed enough that, by the first day of 1948, the nation had a new Educational Testing Service.

Since its founding in 1900, the College Board had consistently played a large role in defining requirements for admission to college; generally the Board acted as a restriction on possible definitions of requirements.¹ Even under the new plan implemented in 1916, curricular latitude for the secondary schools was limited. This restrictiveness was, in part, responsible for a pronounced decline in students taking Board regular examinations between 1929, when 23,478

¹ John W. Valentine, The College Board and the School Curriculum (New York: College Entrance Examination Board, 1987), pp. 25-30; See also "Report of the Associate Secretary" in College Entrance Examination Board, Thirty-Third Annual Report of the Secretary of the College Board (New York: College Board 1933), p. 13.

students took the tests, and 1933 when the number was only 17,695.² Simultaneously, the number of students taking the SAT grew, if modestly, from 8,645 in 1929 to 8,932 in 1933.³

More significant than the numbers themselves, was an increasing practice of colleges requiring only the SAT. By 1941 virtually all of the College Board examinees took the SAT. The test had gained the confidence of the Board to the point where John M. Stalnaker could propose and have accepted a plan to use the SAT as a benchmark for equating all College Board Achievement Tests. The mean and standard deviation of any particular achievement test was henceforth equated with the mean and standard deviation of scores of all candidates taking that particular exam and the same group's scores on the SAT.⁴

In 1943 Brigham died, leaving his inchoate ideas regarding a new science of education and his opposition to a national testing agency without a spokesperson. Thus, when William Learned and Ben Wood promoted their consolidation plan after World War II, the strident voice that had served to raise crucial issues about the consolidation was silent.

Jenne Britell, former Director of Information Services at Educational Testing Service, notes in her doctoral dissertation that testing following World

² For the 1930 figure, see College Entrance Examination Board, Thirtieth Annual Report of the Secretary of the College Entrance Examination Board (New York: College Board, 1930), p. 31. For the 1933 figure and a chart illustrating the decline after over a decade of growth, see: College Board, Thirty-Third Annual Report of the Secretary of the College Board, p. 25.

³ For figures on both years, see: College Board, Thirty-Third annual Report of the Secretary, p. 24.

⁴ Wilks, Samuel S., Scaling and Equating College Board Tests: Report on a Study made with the Collaboration of a Study Group of the Educational Testing Service, September 8, 1961. E.T.S. Archives, p. 10.

War II became increasingly detached from pedagogy.⁵ She asserts that the leaders who worked to found ETS focused almost exclusively on "the usefulness of testing in essentially administrative decisions, the decisions for which they were responsible and for which they would be accountable." Before World War I, educational testing was the province of educators; after World War II, "educational testing--as represented by ETS' founding--would become the province of educational policy makers."⁶ Brigham's fear had been that these policy makers who would dominate testing would proceed from an interest in marketing as opposed to an interest in the "the science of education." Whether ETS, chartered on December 19, 1947, represents such dominance of marketing over research is, of course, debatable. Britell contends that "with the establishment of ETS, the founding organizations acknowledged the complexity of standardized testing . . . they also now understood the magnitude of the undertaking required for good test research and development."⁷ Others have argued that the "not-for profit" corporation reflects corporate self-interest and marketing impulses that parallel those that Brigham predicted.⁸

⁵ Britell, "Never Quite a Public Dialogue: The Discussions of Testing in American Education 1897-1964." (Ph.D. Dissertation, History of Education, Teachers College, Columbia University. 1980), 321-325.

⁶ Ibid., pp. 354-355.

⁷ Ibid., p. 354. ⁸ See for example, Allen Nairn, The Reign of ETS The Corporation That Makes up Minds (Washington, D.C.: Ralph Nader, 1980), pp.28-55; David Owen, None of the Above: Behind the Myth of Scholastic Aptitude (Boston: Houghton Mifflin Company, 1985), pp. 1-4; and James Crouse and Dale Trusheim, The Case Against the SAT (Chicago: University of Chicago Press, 1988. pp. 1-3.

1941: A Pivotal Year for the SAT

The year 1941 was a turning point not only for the country at large; it was also a crucial year in the development of the Scholastic Aptitude Test. In 1941 the College Board decided that all forms of the SAT should be equated, so it initiated a complicated procedure to assure that the test remained essentially the same from one year to the next. Also that year, immediately after the bombing of Pearl Harbor, the Board decided to drop the essay form of its examinations and to rely exclusively on its flagship product, the SAT. The war that America entered in 1941, set in motion demographic changes that, immediately after the armistice, magnified the importance of the Scholastic Aptitude Test.

Equating the SAT

The original SAT committee was explicit in conveying to the Colleges that the tests were not going to "attempt to measure scholastic aptitude in definite units constant from year to year."⁹ In contrast, the test makers today contend that the meaning behind particular scores does not change.¹⁰ From 1926 through 1940, Brigham and his research associates in Princeton reported scores

⁹ Carl Campbell Brigham, et. al. "Scholastic Aptitude Tests: A Manual for the Use of Schools," Prepared by the College Entrance Examination Board, p.19. [Document is housed in Educational Testing Services Archives.]

¹⁰ This is a potentially controversial contention. In large part, the testing experts have been able to successfully argue that a change in national average scores indicates a change in the academic potential from one cohort of test-takers to another. For example, James R. Flynn, in a study in the Psychological Bulletin, notes that people today who take I.Q. tests from the 1950s outscore the earlier generation on those tests. He indicates that this demonstrates a growth in basic cognitive skills. Further, he argues that the downward trend, of SAT scores over several recent decades, reflected a worsening of high school education. Discussed in Daniel Goleman, "An Emerging Theory on Blacks' I.Q. Scores" New York Times Education Life, (April 10, 1988):23. This assumption that aggregate test score decline indicated weakness in the education system or conversely that test score increases indicated emerging strength in the system is based on the implicit belief that the test itself does not change or drift. Psychologists have spent much professional energy promulgating that belief.

on the SAT on a scale for which they set each year.¹¹ The tests were not equated from year to year; thus the testmakers had latitude to innovate.

This ambiguity in the relationship among scores from different years did not please, however, the admissions officers. In 1941 the Board responded to pressure from admissions officers and began a process of formally "equating" all forms of the SAT from one year to the next. This equating process necessarily restricts changes in the test. Thus, in contrast to Brigham's eclectic experimentation with early versions of the SAT, versions since 1941 have been stable. ETS researcher Samuel Wilks reflects the change in attitude in contending that, if the format, item types, or score report format of the SAT were to change, "the confusion which would be created among present College Board users who had already developed a familiarity and working facility with the existing scale would be intolerable."¹²

In 1961 Wilks reported to ETS President Chauncey that "the scales used for all forms of the SAT which have been developed since 1941 are linked directly or indirectly to that used for the April 1941 form of the SAT."¹³ A pattern of linking scores from a previously administered examination produces a complicated interrelationship between forms of the test; they always, however, must purport to measure the same attribute in the same way.¹⁴ According to Wilks:

¹¹ Wilks, Scaling, p. 102.

¹² Ibid., p. 13.

¹³ Ibid., p. 12.

¹⁴ William B. Schrader "Geneological Charts of Form-to-Form Equating of College Board Tests," pp. 30-45. in Wilks, Scaling, p. 32.

In an attempt to keep all of these SAT forms on a common scale, within years as well as between years, the verbal scores of all forms of the SAT are linked directly or indirectly back to the April 1941 SAT-Verbal form through sections of common material. Mathematical scores on the SAT are linked back to the April 1942 form through common items.¹⁵

Linking scores raises several significant questions for the testers. If a system of linking and equating tests could be executed without error, and if candidate performance on the equating material were perfectly stable over time, then the present SAT scale would provide a normative interpretation in terms of the original equating group. As recognized even by researchers for the testing agencies themselves, "the basic question which arises here is how useful at the present time is the equating group as a reference population. . . . Undoubtedly this reference population holds hardly more than an historical interest to any group of SAT users today."¹⁶

The process of equating is complicated and couched in technical language. The test developers discuss a section's genealogy with past sections in terms of "scaling," "stabilization of raw score distributions" and "master reference populations."¹⁷ The process is designed and portrayed in terms meant to inspire confidence among educators that the meaning of SAT scores is consistent from year to year. "College Admission officers need assurance that trends in mean SAT scores reliably reflect trends in academic quality of admitted students."¹⁸ ETS demonstrates concern when evidence of score drift appears.¹⁹

¹⁵ Wilks, Scaling, p. 6.

¹⁶ Ibid., p. 13.

¹⁷ Ibid., p. v. For an indication of the complexity of the mathematics involved in scaling the SAT, see William H. Angoff, "Basic Equations in Scaling and Equating," Appendix 9 in Wilks, Scaling, pp. 121-129.

¹⁸ Wilks, Scaling, p. 14.

The process of equating has, in fact, convinced American educators that the test remains constant and that score changes actually reflect changes in the intellectual or academic aptitude of American students. Throughout the period from the late 1960s through the early 1980s, America faced aggregate score declines on the SAT and focused far more on how education and society had caused a loss of academic aptitude among its youth than on whether the SAT was actually discovering such aptitude.

1941: The Demise of the Essay

Throughout the 1930s discussions about the relative merits of objective examinations as compared to essay examinations were widespread in American education and became a focal point of debates about the College Board Examinations.²⁰ Initially Carl Campbell Brigham defended the essay format against the criticisms of Ben Wood.²¹ The link between the objective testing movement and the proposal for a national testing agency led to Brigham's position. He had no strong objections to objective tests; in fact, in general he was sympathetic to the idea that multiple choice examinations could be

¹⁹ See, for example: S.A. Kendrick, "When SAT Scores go Down," College Board Review 64 (Summer 1967):5-11. Kendrick, a College Board Staff member noted that test score changes, particularly decreases "are a serious business . . . some years are worse than others and 1966-67 was very bad indeed." 1966 marked the beginning of a long and troublesome national score decline.

²⁰ For an important compilation of the issues in the debate at mid-decade, see: "Report of the special committee on Problem:.. American Council on Education." Ben D. Wood Papers MSS 11, ETS Archives.

²¹ Ben D. Wood "Oral History" with David Hubin, Croton-on-Hudson New York, May 23, 1985. Tape and Transcript on file at E.T.S. Archives. pp. 7-9.

developed to test more than the lowest levels of cognition and learning--recall and recognition.²²

Opinions of the Board's older program of examinations were quite mixed; some secondary educators objected to the influence that they had on secondary curricula, while others valued the influence because of its emphasis on writing and synthesis.²³ Harvard's Conant objected to the use of the term essay:

I must quarrel with the terminology of those who have labeled the examinations of the first decades of this century 'essay examinations' . . . The significant characteristics of those earlier examinations was not the long essay-type answer required in some instances. The hallmark was their relation to a carefully prescribed area of content."²⁴

Testers in the thirties were already quite critical of the essay format examinations for any purpose. Wood recalls that there were obvious weaknesses in the grading of the essay examinations.

I used to know some of the readers that read the College Board Examination and some of them became quite candid with me. They would frequently be sitting at the same table as a friend, a close friend, and one of them would score the written test an A and the other one a D. Sitting next to each other, they would swap papers and see these differences.²⁵

The essay examination would have been on its way out even without manpower shortages caused by the War. In the late thirties as they formulated the new Graduate Records Examination, Ben Wood and William Learned consistently criticized the essay format. In the original memo to students who

²² Brigham worked on subject area multiple choice examinations in the late thirties.

²³ Valentine, The College Board, p.52.

²⁴ James B. Conant, My Several Lives: Memoirs of a Social Inventor, (New York: Harper and Row, 1970), p. 421.

²⁵ Wood "Oral History" with Hubin, p. 6.

would be taking the Graduate Records Examination, Learned noted, "The questions, instead of requiring discursive answers, will be of a sort to test knowledge, judgment, and reasoning ability through simple recognition. By this method a large amount of ground can be covered in a short time."²⁶ In January, 1933 William Learned observed that the essay examination "has been almost exclusively a subjective examination, that is, a few arbitrarily chosen questions with answers rated in terms of personal opinion by individuals of varying experience and competence." While commending the College Board for the quality of its work with essay examinations, he argued that "the defects are inherent." "It is because of this notable achievement of the Board in pushing the discursive written examination to the extreme limit of its technical possibilities that the inevitable shortcomings of this type of test have been clearly disclosed."²⁷ Albert Crawford, Director of Admissions at Yale, and John M. Stalnaker of Harvard, both wrote to Learned in February and March of 1938 suggesting that Learned's explanations of the new Graduate Record Examination sent to the students taking the examination in 1937 might have actually gone a little too far. Both, apparently independent of each other, objected to Learned's phrase "unlike the usual type of examination, the Graduate Record Examination is exclusively a test of knowledge." Crawford noted:

I presume that this was merely intended to explain why the G.R.E. was of a different form from the usual examinations . . . As it stands, it seems to imply that the usual examination is not intended to be a measure of knowledge. While some professors contend that the essay examination measures other mysterious powers as

²⁶ Vera S. Fueslein (Secretary to William S. Learned) to Ben D. Wood. May 25, 1937. GRE Papers, Box 1, E.T.S. Archives. [Draft of Learned memo to graduate students]. See also: Hubin, David R., Oral History Interview with Ben D. Wood.

²⁷ William S. Learned, "Admission to College," Educational Record, (January 1933):31.

well as knowledge, I don't think any of them feels that it fails to do the latter.²⁸

More directly, Stalnaker wrote, "I dislike any statement . . . which casts aspersion, even by indirection, on any other examination form."²⁹ Stalnaker's seven page letter severely criticized Learned for claiming too much for the new examination and for its implicit criticism of other examinations, including the essay examinations. He concluded by arguing that "obviously we have much work to do before the Graduate Record Examination is an instrument in which we can place considerable faith . . . In the meantime, is it not the policy of wisdom to content ourselves with claiming only that of which we are certain, even though our claims be meager. I wish again to express my distress that the statement herein criticized has already been given some circulation."³⁰

Despite his rejoinder to Learned, Chauncey was by no means a consistent defender of the essay format examinations, particularly when these tests were compared to the newer SAT. In retrospect he stresses his criticism of the examinations and recalls an example of how the old essay examinations were, in his words, "unfair to students who went to schools that did not prepare for them [for particular content exams]". He cites the case of a student from Milwaukee who was "fourth or fifth in a class of five or six hundred students," and who was promised a scholarship. The student failed all four of the essay examinations but did well on the SAT. Chauncey advocated this boy's admission "just as an

²⁸ Albert Beecher Crawford to William S. Learned. February 21, 1938. G.R.E. Papers, Box 1. E.T.S. Archives.

²⁹ John M. Stalnaker to William S. Learned. March 2, 1938. G.R.E. Papers, Box 1. E.T.S. Archives. p.1.

³⁰ Ibid., p. 7

experiment, [to] try him out and see what happens." It turns out that in this anecdotal example, the Milwaukee youth became an A and B student.³¹

In order for the Board to drop the essay examination, two factors were necessary. First, the advocates of objective types of examinations needed an efficient way to score the new tests; second the leaders of the Board, once convinced of the value of making a change, needed a justification to mollify the supporters of the essay. Both of these factors were present in 1941.

The Introduction of Machine Scoring

By the end of the 1930s, many educators realized that inefficiencies in scoring examinations limited the future of large scale testing programs. Hand scoring of test booklets, whether they represented traditional essay examinations or free response short answer items, was tremendously slow and time consuming. Ben Wood first became concerned with this inefficiency when he worked with the New York Regents examinations; he faced it again with the Pennsylvania study.³² In late 1928 Wood wrote to ten corporations that manufactured business equipment. In response, Thomas J. Watson telephoned Wood and asked for a meeting. Subsequently, Watson provided Wood with three truckloads of IBM computing equipment and hired the psychologist as consultant on a project to develop test scoring machines.³³

³¹ Chauncey, Henry, with Gary D. Saretzky. Oral History with Henry Chauncey, March 28, 1977. ETS Archives Oral History Program. Educational Testing Service, Princeton New Jersey. March 31, 1977, p. II-6.

³² Matthew P. Downey. Ben D. Wood: Educational Reformer (Princeton, N.J.: Educational Testing Service, 1965), p. 49. Downey provides a useful discussion of the development of machine scoring and of Wood's long relationship with Thomas B. Watson. See also Wood "Oral History" with Hubin, p. 16.

³³ Downey, Wood, p. 51.

Wood experimented with various versions of tabulators; all of them, however, proved too slow until he employed a concept of electric conductivity through pencil lead that had been developed independently by Reynald B. Johnson, a Michigan high school science teacher. In 1934 IBM hired Johnson to head the development process for the machine that Wood so actively sought.³⁴

A workable efficient mechanized approach to scoring examinations was a major breakthrough for those who favored broader use of tests. As Arthur Traxler of the Educational Records Bureau noted in 1953, "If the Model-T put America on wheels, the test scoring machine put the youth of America on objective-test answer sheets."³⁵ Traxler also raised a note of concern in 1953; he feared that the introduction of machine scoring contributed to a process of fixing the examinations within a particular format, saying that "the fixed response position and the fixed fields of the scoring machine has tended to force objective testing into a kind of strait jacket."³⁶ Certainly the advent of machine scoring contributed to the ascendancy of the SAT; the machine further tipped the cost/efficiency ratio between the old-style essay examination and the multiple choice aptitude test in favor of the latter.

In his 1941 report as Executive Secretary, George W. Mullins noted that "we have made a wider use this year of the machine system of recording and

³⁴ Ibid., p. 53

³⁵ Arthur Traxler "The IBM Scoring Machine: An Evaluation," 1953 Invitational Conference on Testing Problems (Princeton, New Jersey: Educational Testing Service, 1953), p. 140. Cited in Downey, Wood, p. 53-54.

³⁶ Ibid. p. 140.

reporting the results of the examinations."³⁷ From that point forward the SAT was entirely machine scored; no items such as free response math problems that could not be corrected by machine ever appeared again on the examination.³⁸

Manpower Shortage: The Rationale for the Essay's Demise

America's entry into World War II gave the leadership of the Board the excuse they needed to drop the essay examination. Henry Dyer reports that "it wasn't until, you know, after Pearl Harbor that colleges went on the three-term year and were admitting students the year around, and the old essay tests had to be dropped from the College Board program."³⁹ According to Dyer, "George Mullins and Dick Gummere and Rad Heermance, sitting over here in Princeton, were discussing it even before Pearl Harbor. It was on that Sunday and they got the call, and decided to drop the essay."⁴⁰

The manpower shortage was the consistent explanation. In response to a 1942 request from Nicholas Murray Butler for information on the activities of the Board, George W. Mullins wrote: "Immediately after Pearl Harbor when the

³⁷ Forty-First Annual Report of the Executive Secretary (New York: College Entrance Examination Board, 1941), p. 3.

³⁸ For a detailed list of the items on the SAT between 1926 and 1960 see Loret, pp. 26-90.

³⁹ Dyer, Henry and Gary D. Saretzky. Oral History with Henry Dyer, September 25, 1978, ETS Archives Oral History Program Educational Testing Service, Princeton New Jersey, 1980. p. 8.

⁴⁰ Ibid.

colleges and universities adopted the all-year around accelerated program, it became evident at once that the only way for the Board to meet promptly and adequately the exigency of the situation was to replace the seven day examination period in June by the scholastic aptitude and scholastic achievement test given on one day in April. This was immediately done." Butler's reaction to the change is uncertain.⁴¹ Mullins stressed, "Since these tests have been perfected by an experimental program of research, the Board was adequately prepared for just such an emergency as confronted it."⁴²

Because the wartime manpower shortage was given as a reason for the demise of the essay examination, those who supported that form of testing naturally wondered whether they would be reintroduced following the war. In the 1942 Annual Report, Executive Secretary George W. Mullins made clear, however, that the "the passing of the June essay-type examination after continuous use for forty-one years marks the end of an era so far as the history of the Board is concerned."⁴³ Henry Chauncey notes "at the end of the War the

⁴¹ Inconclusive evidence presented in an oral history of Ben D. Wood indicates that Butler was antagonistic toward the SAT; Wood recalled: "the only time I ever saw him turn red with anger was when he read that one of those outrageous things that Brigham or...He said no man of his kind of a mind can write a good examination or anything else good." The span of time between the events and the oral history, and Ben D. Wood's own critical assessment of Brigham must be considered in reaching conclusions. Wood "Oral History" with Hubin, p. 15.

⁴² George W. Mullins to Nicholas Murray Butler. August 1, 1942. College Board Archives.

⁴³ Forty-Second Annual Report of the Executive Secretary (New York: College Entrance Examination Board, 1942), p. 3. Mullins and others expressed the same thoughts less publicly. In writing to Myra McLean, Assistant Secretary to the College Board, Mullins noted, regarding any possible return to essay examinations: "certainly not for the duration of the war, probably never. This is what I hear on all sides." G.W. Mullins to Myra McLean July 25, 1942. College Board Archives..

colleges were so satisfied with this testing program that they decided not to revive the old essay examinations."⁴⁴

After the termination of the essay examinations, the Board made a concession to the secondary schools that had supported the tests; examination questions in English and American history were made available to any secondary school that wished to administer and grade them as a form of "terminal school examination" but not for use in any way for college admissions.⁴⁵

Testing for the Military

Among psychologists and educators during World War II, military and personnel issues overshadowed considerations related to the SAT specifically or to college admissions generally. Unlike World War I, during which "the usual program of the Board was undisturbed,"⁴⁶ World War II dramatically changed the work of the Board.⁴⁷ President Seymour of Yale referred to the colleges during World War II as "an educational task force," and the Board's executive secretary George W. Mullins noted that "the most important concern of schools and colleges is their contribution to the war effort of the nation."⁴⁸ The College Board helped marshal the resources of higher education for the wartime effort.

The war and the military imperative for instruments to place individuals into appropriate roles enhanced further the role of psychologists within the College

⁴⁴ Henry Chauncey to Mrs. Sharp, September 27, 1961. Brigham Papers, ETS Archives.

⁴⁵ Valentine, The College Board, p.53.

⁴⁶ Forty-Second Annual Report of the Executive Secretary, p. 3.

⁴⁷ Valentine, The College Board, p. 51.

⁴⁸ Seymour quoted in Forty-Second Annual Report of the Executive Secretary, p. 4.

Board. John Valentine points out that the activities that might be called "Board" activities during the war were actually the work of a growing number of psychologists and test development specialists "based in Princeton whose activities and feats were baffling to most member of the Executive Committee and to many representatives of member colleges."⁴⁹

Harold Gulliksen, a psychologist who had worked with Herbert Toops at Ohio State and with Louis Thurstone at Chicago, was among the psychologists drawn to Princeton. Gulliksen, who would assume the role of research secretary for the Board after Brigham's death in 1943, took a leading role in the organizations testing research and development for the military. Gulliksen and fellow psychologist Norman Fredericksen abandoned or questioned old testing paradigms as they developed instruments to measure competencies ranging from general mathematical skills to the mechanical ability to assemble a weapon.⁵⁰

Thus the psychologists at the Princeton offices of the Board established a foundation for practical vocational testing that would later be a part of the work of Educational Testing Service. The psychologists tested for requisite skills to accomplish particular tasks in the war effort; their primary emphasis was thus not to rank candidates but rather to ascertain whether an individual met a criterion of successful performance. In these efforts psychologists were creative and

⁴⁹ Valentine, The College Board, p. 53.

⁵⁰ Gary D. Saretzky, "Interview with Harold O. Gulliksen: March 3, 1975" ETS Oral History Program, ETS Archives, p.7. The Board produced 133 different tests for the Navy alone. see: Valentine, The College Board, p. 53.

innovative and reflected the advances in learning theory.⁵¹ These experimental links between pedagogy and testing stood in clear contrast to the of the static SAT--an instrument that continued to simply arrange candidates along a standard bell curve.

SAT During the War

The war demonstrated and even produced flexibility in measurement generally, and the war changed the environment in which the Scholastic Aptitude Test operated. However, the war's effect on that examination itself only involved the Board's scheduling earlier administrations of the test.⁵² Because special War department programs on college campuses began in June, it was not possible to use the June tests. On December 22, 1941, representatives of Barnard, Bryn Mawr, Mount Holyoke, Radcliffe, Smith, Vassar, and Wellesley met in the New York offices of the Board and voted to "direct their final candidates for admission to take the Scholastic Aptitude Test and the Achievement Tests to be held in April."⁵³ Although Board leaders objected to the change because of the impact that testing during the academic year might have on secondary school curricula, "it was pointed out that the Board was, of course, dependent upon the colleges and could give only such examinations as the colleges found useful."⁵⁴ The Board began discussion of a December examination date.

⁵¹ See Gulliksen's specific discussion of competency testing and learning stations for gunnery specialists. *Ibid.* pp. 10-13.

⁵² John M. Stalaker to Mrs. Ralph J. Sharp, September 28, 1961. ETS Archives.

⁵³ Minutes of the College Entrance Examination Board Joint Meeting of the Advisory Committee on Problems and Policies and the Committee on Finance, April 7, 1942. Page 5. College Board Archives.

⁵⁴ "Minutes of the College Entrance Examination Board," April 7, 1942, p. 5. College Board Archives.

The Postwar Explosion of Testing

The national mobilization necessary during World War II had focused the attention of psychology on issues directly related to the military action. At the end of the conflict, some researchers continued to study the traits identified in the military research; others returned to investigation of the meaning of intelligence and the development of intelligence tests.⁵⁵

In 1947 Warren G. Findley, William W. Turnbull, and Herbert S. Conrad called for a change in emphasis of the research on intelligence testing. Psychologists had accepted the tests; now the forward looking among them were looking at refinement but only within established boundaries. The issues these researchers raised were essentially second generation questions. Noting that research to that point had focused primarily on the correlation of one test with another, the reliability of various instruments, and the correlations of the tests with performance criteria such as grades, these psychologists invited research on other problems of test construction, such as appropriate techniques for item writing, proper item mix for maximum validity, the effect of coaching on test performance, and the effect of speededness of tests on validity.⁵⁶

As the Second World War ended, the College Board made available a special test designed for veterans who had been away from formal education for several years. Administered first in November 1944, the Special Aptitude Test for Veterans consisted of three forms, each containing a verbal and a

⁵⁵ See, for example: Carnegie Foundation for the Advancement of Teaching, "An Inquiry in Postwar Conditions in American Colleges: A Proposal to Sample Progressively by Examination The Current Academic Equipment of College Sophomores and Seniors," ETS Archives, Microfilm File.

⁵⁶ Warren G. Findley, William W. Turnbull, and Herbert S. Conrad, "Construction, Evaluation, and Applications of Intelligence Tests," Review of Educational Research 17 (February 1947):13.

mathematical section patterned after the SAT. Candidates could then select either a "spatial," a "science" or a "social studies" test. The spatial items paralleled sections that Brigham had experimented with but not introduced for the SAT itself. The other two sections tested knowledge and reading ability in the content areas of either science or social science. Initially the Board simply made the test available to member colleges for on-campus administration. In September 1945 the SATV became an option at the regular quarterly Board examinations, and in January 1946 the Board authorized bi-weekly administrations in nine cities: Berkeley, Buffalo, Cambridge, Evanston, Los Angeles, New York, Philadelphia, Pittsburgh, and Washington, D.C. In 1948 by the end of the special program, 35,742 veterans had taken this special examinations.⁵⁷ Henry Chauncey took a leading role in introducing the Special Aptitude Test for Veterans, his statements indicate his vision for an emerging role for the Board. "I believe that the Board's future reputation depends on its assuming leadership on the frontiers of testing. This does not mean that the Board should publish or administer tests without adequate experimentation and full proof that they are useful instruments, but it does mean that the Board has to venture forth."⁵⁸

Just as postwar demographic trends increased the importance of the SAT, so too did these trends enhance the perceived importance of the Graduate Record Examination. The impulse to broaden the use of the Graduate Record Examination was strong. As Stuart Peterson noted in a 1943 doctoral

⁵⁷ "Statistical Summaries" SATV File, Folder 1. Educational Testing Services, Princeton, New Jersey.

⁵⁸ Henry Chauncey to John Stalnaker, July 25, 1944. Educational Testing Service Archives, SATV File, Folder 1. ETS Princeton New Jersey.

dissertation, "The very rapid increase in the number of graduate students registering in the graduate colleges of the United States within recent years has presented administrators and faculties with serious problems. . . . The need for a scientific appraisal of entering graduate students is obvious."⁵⁹ On March 2, 1944, after what the Carnegie foundation referred to as a five-year experimental period, the Graduate Record Examination became available to "any student desiring to present the results in connection with his admission to American graduate or professional schools."⁶⁰ In announcing the availability of the test, the Carnegie Foundation asserted that "studies of the results at several universities have shown that the examination alone predicts success in graduate school about as well as the undergraduate record."⁶¹

The G.I. Bill and American Colleges

Immediately following the armistice after World War II, the number of prospective students seeking admission to higher education jumped dramatically.⁶² As Algo Henderson puts it, "the generation that had graduated from high school wanted its children to attend college. Following World War II, the wave swept into the colleges."⁶³ This wave, to an unprecedented degree,

⁵⁹ Peterson, Stuart Conrad, "The Measurement and Prediction of Scholastic Achievement on the Graduate Level," Ph.D. Dissertation, Department of Psychology, the State University of Iowa. 1943. Published by the Graduate Record Office p. 1.

⁶⁰ "A General Examination For Advanced College Students," ETS Archives, p. 1.

⁶¹ "A General Examination For Advanced College Students," ETS Archives, (Check Microfilm Number on E.T.S. Microfiche.) p. 2.

⁶² CFAT, "An Inquiry into Postwar Conditions," p. 1.

⁶³ Algo Henderson and Jean Glidden Henderson, Higher Education in America: Problems, Priorities and Prospects (San Francisco: Jossey Bass, 1974), p. 3.

found financial support from the federal government.⁶⁴ One of the forms of support was through the Servicemen's Readjustment Act of 1944--the so called G.I. Bill of Rights. In the fall of 1946 over a million war veterans, with government financial support, swelled the enrollments of American colleges; in one year the number of college students doubled.⁶⁵ Although support for veterans did not signal a commitment to a principle of the federal government providing aid to all deserving college students, it did set a precedent that extended financial aid programs.⁶⁶

Along with the growth of college enrollments the Board's testing role grew. So too, did the Board's impact on college admissions grow. However, the Board in the mid 1940s also participated in the formation of a new organization that would create for it what amounted to an identity crisis.⁶⁷ The College Board's group of psychologists in Princeton, a group that had grown in number and importance during the war, would soon form the core of a national testing agency.

Educational Testing Service: The Final Ossification

Shortly after the war, William Learned once again made overtures to the College Board about the possibility of combining the Scholastic Aptitude Test, the G.R.E., the Cooperative Tests Service, and The Educational Records Bureau into a new testing agency. He sought merger of three separate testing agencies "in

⁶⁴ Brubacher, Higher Education in Transition, p. 233.

⁶⁵ Valentine, The College Board, p. 65.

⁶⁶ Ibid., p. 230.

⁶⁷ Frank Bowles, The Refounding of the College Board, 1948-1963: An Informal Commentary and Selected Papers (New York: College Entrance Examination Board, 1967)

order to avoid overlapping and unproductive competition and in order to make more resources for research and development."⁶⁸

Learned's initial plans simply called for the College Board to take over the administration of the G.R.E. He was about to retire from the Carnegie Foundation and "he needed some place to lodge the Graduate Record Examinations and that would take care of his 'baby.'"⁶⁹ This plan met resistance; word got to the American Council on Education and to many G.R.E. users who were not members of the College Board; these groups were "kind of unhappy at this potential development."⁷⁰ The College Board, with its membership of approximately fifty colleges, was at this time, in the words of Henry Chauncey, "a very elite organization, particularly in the eyes of ACE," primarily focused on the Ivy League.⁷¹

Bowing to this pressure not simply to give the Graduate Record Examination to the Board, the Carnegie Foundation appointed James B. Conant to head a commission to look for a home for the G.R.E. Conant had become

⁶⁸ Conant, My Several Lives, p. 417. See also Frank Bowles, The Refounding of the College Board, pp. 1-4.

⁶⁹ Chauncey, Oral History, April 15, 1977. p. IV-2. E.T.S. Archives. Chauncey refers to Learned as "in a certain sense, he was the initiator or the stimulator." He attributes Conant, however, with being the "intellectual godfather" of E.T.S. "Conant and the people behind the scenes. . . were the ones who helped to navigate through Scylla and Charybdis so something happened." pp. IV-2. Chauncey later uses a slightly different metaphor; he indicates that the Carnegie Foundation wanted to turn the G.R.E. "over lock, stock and barrel, maybe with a little dowry because it had obviously been supported by the Foundation and it wasn't a self-supporting activity, to the College Board." pp. IV-6.

⁷⁰ Chauncey, "Oral History," April 15, 1977, pp. IV-3.

⁷¹ Ibid., pp. IV-4,5. The Chauncey interview presents significant points on the attitude of such men as Sproul of the University of California, a man actively involved in the American Council on Education, toward the College Board.

enamored with the SAT during the 1930s.⁷² He felt that this test should be central to a new testing agency, but, given the reaction of the American Council of Education, he did not seek to simply enlarge the role of the existing Board. When the Conant Committee met, it recommended, rather, the formation of an entirely new organization, the General Examination Board. Learned discussed the matter of consolidation again with George Mullins, secretary to the College Board and indicated that "he was retiring soon and that it was the desire of the Carnegie Foundation and Carnegie Corporation to terminate the operation of the Graduate Record Examination and the Pre-Engineering Inventory Tests and yet to find some way to have these tests carried on independently of the Corporation and the Foundation."⁷³

The Carnegie Foundation played a crucial role in bringing together the components of a large nationally recognized Educational Testing Service."⁷⁴ First, the Carnegie Foundation was instrumental in persuading the American Council on Education to turn over its testing functions--most visibly the Psychological Examination--to the new organization. Second, the Carnegie Corporation agreed to make them a grant of \$50,000 a year; that "naturally put a

⁷² Spring, Joel, The Sorting Machine: National Educational Policy Since 1945 (New York: Longman, 1976), p. 43. John W. Valentine, too stresses the significance of Conant's interest and that of his staff, Richard M. Gummere and Henry Chauncey. Valentine, The College Board, p. 48. See also Ellen Lagemann, Private Power for the Public Good: A History of the Carnegie Foundation for the Advancement of Teaching (Middletown, Connecticut: Wesleyan University Press, 1983), p. 115.

⁷³ Statement from George Mullins, quoted in James B. Conant, My Several Lives, p. 429.

⁷⁴ Ibid.

little sugar coating on the arrangement."⁷⁵ The Carnegie Foundation's direct involvement continued through the early years of ETS. Then, "Carnegie decided after a number of years that the President of Carnegie no longer needed to be a member" of the ETS Board of Trustees.⁷⁶

The final proposal from the Conant Committee called for the Board to continue as a "consumer organization,"⁷⁷ with its testing and test development activities contracted out to a new national testing agency. The State of New York formally granted a charter to the new Educational Testing Service on December 19, 1947. According to the Board's announcement of the formation of E.T.S. "the College Board as an association will lose nothing. Rather, it will gain, by being better able than it has been in recent years to fulfill its original purpose of being a deliberative body devoted to problems of college admission. Specifically, the Board will gain by being relieved of the management of an operating office and of concern for various testing activities recently undertaken for the government and for other organizations."⁷⁸

With Brigham gone, so too was the most vocal and visible opponent of a national testing agency. Brigham, however, was not the only researcher who had

⁷⁵ Chauncey, Oral History. April 15, 1977, pp. IV-14. See also: Henry Dyer and Gary D. Saretzky. "Oral History with Henry Dyer," September 25, 1978, ETS Archives Oral History Program Educational Testing Service, Princeton New Jersey, 1980. p. 20.

⁷⁶ Chauncey, "Oral History," April 15, 1977, pp. IV-24.

⁷⁷ Ibid., p. IV-10. For a discussion of the impact that this change had upon the College Board see Valentine, *The College Board*, pp. 63-67; for the perspective of the first Board President following ETS's formation and for selected documents see: Bowles, *The Refounding of the College Board*, pp. 1-48.

⁷⁸ "Henry Chauncey and Edward S. Noyes, "Announcement: Merger of Testing Offices with Those of the American Council on Education and the Carnegie Foundation" December 18, 1947. College Board Archives.

concerns about such an agency. Henry Dyer, later an ETS Vice President, recalls in a 1978 oral history that initially he had "grave doubts" that led him to write a confidential memorandum to his boss at Harvard. His recollection was that he said "Because I think that if you set up an agency of this sort, it will lose touch with the grass roots, which is what the Board is capable of doing because of its membership and so on. And this could be serious."⁷⁹ Dyer's confidential memo was passed along to President Conant, the strong advocate of the mergers, but it had no effect.

The Success Story

Although creation of Educational Testing Service created an initial identity crisis for the College Board, the Board soon redefined itself and, in concert with its progeny, ETS, expanded its role in college admissions. In the decade after ETS's formation, the Board became a truly national organization. Whereas in 1947 over seventy-seven percent of the Board's total membership of sixty-seven institutions came from New England, by 1959 that percentage had declined to fifty percent of a total membership of 286 institutions.⁸⁰

⁷⁹ Dyer, Henry and Gary D. Saretzky. Oral History with Henry Dyer, September 25, 1978, ETS Archives Oral History Program Educational Testing Service, Princeton New Jersey, 1980. p. 21.

⁸⁰ The membership changes were as follows:

		1947		1959
New England	(52)	77.5%	(145)	50.5%
South	(5)	7.5%	(67)	23.4%
Midwest	(4)	6.0%	(44)	15.3%
Mountain and Pacific	(6)	9.0%	(30)	10.5%

Henry Dyer and William Turnbull, "Growth and Change in College Admissions Testing" (Thirty-one page memorandum originally labelled as "confidential."), p. 7. On Microfilm at ETS Archives.

Conclusion: Trends Continue to the Present

The formation of Educational Testing Service established a foundation for testing to become a large business. In the decade of the fifties, no clear competition in admissions testing existed for the new organization. Expansion of admissions testing involved convincing additional institutions to adopt the process. Such selling was not always easy. For example, in the West, recalls former College Board executive vice president Richard Pearson, "the big problem, in a competitive sense, was the University of California, which dominated the California higher education scene. In those days, they were quite selective, but without the use of tests."⁸¹ This outstanding state system of universities based admissions on "high school standing, pure and simple."⁸²

Reluctance of such a prominent university notwithstanding, the Educational Testing Services and its flagship product, the SAT, grew dramatically throughout the 1950s. Over 81,100 students took the SAT in 1951; ten years later, in 1961, the number had increased ten-fold to 805,500.⁸³

As the numbers of students taking the SAT changed, the test itself remained constant. Its static nature is a realization of Carl Campbell Brigham's worst fears. In 1937 he had predicted that wide use of the examination, with large samples and established norms, would lead to the test's stagnancy. He foresaw and feared that marketing would overshadow experimentation and science in testing. Warren G. Findley, Brigham's successor, reveals that the

⁸¹ Richard Pearson and Gary Saretzky, "Oral History with Richard Pearson," ETS Archives Oral History Program December 27, 1979. ETS Archives, p. 10.

⁸² Ibid.

⁸³ College Entrance Examination Board, On Further Examination: Report of the Advisory Panel on the Scholastic Aptitude Test Score Decline (New York: College Board, 1977), p. 4.

Scholastic Aptitude Test was such a remarkable success from the start that there was little interest in Brigham's groundbreaking work on determining why people make errors.⁸⁴

Between 1926 and 1941 there had been a good deal of experimentation on content and format of items in both the verbal and mathematical sections of the SAT. But since 1941 there have been relatively few changes."⁸⁵ The SAT reflects very directly its early heritage and emerges relatively uninfluenced by groundbreaking work in learning theory and cognitive psychology. ETS researcher William Coffman allows that "an examination of successive forms of the SAT may suggest that the changes subsequent to the original period of development have been minor."⁸⁶ He then defends the consistency by noting that "it is this comparative stability of the scale which has made possible the accumulation of data which enriches the context within which individual scores are interpreted."⁸⁷ As an instrument is proposed to meet a perceived need, creativity is a dominant characteristic; however, when the instrument appears to meet that need, further creativity becomes circumscribed, troublesome to the marketers who take over from the researchers.

⁸⁴ Findley, Warren G. "Carl C. Brigham Revisited" The College Board Review 119 (Spring 1981):8

⁸⁵ Wilks, Samuel S., Scaling and equating College Board Tests: Report on a Study made with the Collaboration of a Study Group of the Educational Testing Service, September 8, 1961. E.T.S. Archives, p. 7. The 195 page Wilks report, commissioned by E.T.S. President Chauncey to address issues in equating scores from one year to the next, covers in detail the genealogy of item types and the changes. This study specifically addresses the question of how "useful the 1941 reference group is for candidates taking the SAT."

⁸⁶ William E. Coffman, "Research and Development Report--The Scholastic Aptitude Test - 1926-1962." Test Development Report TDR-63-2, June 1963. E.T.S. Archives, p. 13.

⁸⁷ Ibid., p. 9.

The SAT truly does represent a "commodity science." Entering its sixty-third year, the SAT continues to occupy a prominent position in the college admissions process. More significantly, despite recent criticisms of the SAT within both the professional and popular media, in large part the American population has accepted the test to be something much more than a measure of an individual's ability to take that particular examination. Educators, parents, and their children have elevated the meaning of the examination beyond what even the researchers who developed the test would claim. With few questions and for over six decades, Americans have accepted the SAT as a valid test of a person's potential intellectual capabilities; perhaps in this ironic light, the SAT has, in fact, measured something important about the American people's intellect.