

## CHAPTER VI

### CONFIGURING AND ADMINISTERING THE FIRST SAT

#### Introduction

Slightly over eight thousand candidates took the first Scholastic Aptitude Test on June 23, 1926. Although twenty three different colleges received the scores of these candidates, the June date and the experimental nature of the examination meant that "in most cases decisions regarding admission or rejection were made without reference to the test scores."<sup>1</sup> Although the test had little if any impact on futures of those first students, in the broader context of college admissions that first Scholastic Aptitude Test was historic. The SAT those students took directly reflected both its heritage from the Army Alpha Test and the particular stamp of Carl Campbell Brigham.

This chapter discusses the technical decisions necessary in configuring the earliest SATs. Further, it treats the early modifications that Brigham introduced after the first two years of testing. To Brigham, the Scholastic Aptitude Test was a tool that should always remain somewhat fluid and should always be based on current research. The College Board leadership, in contrast, ultimately sought a codified instrument that could be marketed.

#### The Novelty of the Examination

As the "Technical Committee of Experts" finished work on the first Scholastic Aptitude Test, the Board prepared to administer the new instrument. On January 9, 1926, Thomas Fiske assembled the group of supervisors for the first examination. Fiske began the meeting by announcing: "The examinations

---

<sup>1</sup> Coffman, William E. "Research and Development Report--The Scholastic Aptitude Test - 1926-1962" Test Development Report TDR-63-2, June 1963.

next June will be characterized by a decided innovation. There will be a psychological examination on Wednesday morning, and according to the communications that I have received, the supervisors are all interested in the character of the psychological examination."<sup>2</sup> Although Fiske was not himself clear on the specific configuration of the new test because "the committee in charge of the psychological examination [had] not completed the construction of the examination," he acknowledged that "the psychological examination will be very different from the other examinations."<sup>3</sup>

Fiske then discussed the new test in terms reflecting the ambivalent and uncertain attitude that permeated the Board. He predicted that "almost all the colleges will, in the end, demand or request their candidates to take the psychological examination."<sup>4</sup> Moreover, he foresaw the examination playing a role in the admissions decision itself, not simply as a post admissions validating tool; he anticipated that most colleges would eventually tell candidates to take the psychological examination in June for consideration in admissions.<sup>5</sup> However, Fiske also candidly told a test administrator that:

We have cleared the deck for the psychological examination, and we have given the psychological examination perhaps the most favorable period, because if the psychological examination should result in disaster, if the experiment should prove to be a failure, we do not want the psychologists to say, 'That is because you held the

---

<sup>2</sup> Minutes of the "Conference of Supervisors Held in the Trustee's Room, Columbia Library, Saturday, January 9, 1926, p. 3. Educational Testing Services Archives. [No adjournment time for the meeting is stated in the minutes, but based on the 133 pages of minutes it was a very thorough discussion of the new examination. The document provides useful insights into the non-psychologists' perceptions of the new examinations.]

<sup>3</sup> Ibid., p. 3.

<sup>4</sup> Ibid.

<sup>5</sup> Ibid., p. 7.

examination at an unfavorable moment, or when the candidates could not get there, had not found out where the examination was,' or something of that sort.<sup>6</sup>

Such timidity regarding the new examination led the Board to circumscribe the early role of the new SAT. Initially it viewed the new instrument as a supplement to the existing examinations. No candidate could take the "new psychological examination" without being registered with the Board and without taking the board's other examinations. Further, the Board made no initial attempt to pressure member institutions to use the examination. Prior to the first administration of the test, Thomas Fiske made clear that the test would only be given to candidates who were required to take it by the institution to which they were applying.

Candidates will be allowed to take the psychological examination only if the colleges to which they are going consider it important or desirable; that is, if any college suggests that it would be a good thing for the boy or girl to take the psychological examination, the candidate will take it; but if the college does not want the result, there is no use in having the candidate take it, and if the boy or girl says 'well, I haven't the slightest idea of what college I am going to, I would like to get credentials from the Board and then send them to some colleges as yet unspecified,' we will say, 'no.' The candidate will not be able to take the psychological examination and peddle it around from one college to another, trying to see which college will interpret it most generously.<sup>7</sup>

Brigham and his group of psychologists recognized the newness of their examination and realized the importance of alleviating initial suspicions. Having completed the work of writing the new test, the group of psychologists faced introducing and explaining it first to the board and then to educators in general. They decided to publish a "manual" on the test, which began with a brief history of testing and of the need for improved methods of admission. The manual noted

---

<sup>6</sup> Ibid., p. 18.

<sup>7</sup> Ibid., p. 14.

that "during the past decade, colleges have shown an increasingly critical interest in methods of admission," because of the "excess of young people desiring a college education, and a genuine interest in rendering greater service to those actually admitted."<sup>8</sup>

Taking a balanced position on the potential uses of the new psychological tests, Brigham's manual pointed to the evidence of increasing numbers of applicants to colleges and to a problem created by "a large number of young people in these schools, who [although] there seems reason to believe are potential college failures, are nevertheless aspirants for a college education."<sup>9</sup> The new test could, according to the manual, help in determining fitness. Brigham and the committee recognized limitations of the test, however, and stressed that they make

no claim that a candidate with a certain number of points of test score is just able to do college work, or that a candidate with a point or so less is unable to do the work. Realizing that fixed passing marks are not warranted by measures as crude as those now in use, they will name no dead line, but will merely urge that a candidate be judged by the total picture obtained from all criteria available for evaluation at the time of admission. *There is no passing mark for the scholastic aptitude test.*<sup>10</sup>

Brigham himself was concerned not only about any problems the newness of the examination would present for Board members and admissions officers; he worried also about the effect that the test's unfamiliarity would have on test takers. Therefore, the psychologist established an unprecedented requirement

---

<sup>8</sup> Carl Campbell Brigham et. al. Scholastic Aptitude Tests: A Manual for the Use of Schools, Prepared by the College Entrance Examination Board, p. 5. [Document is housed in Educational Testing Services Archives.]

<sup>9</sup> Ibid., p. 6.

<sup>10</sup> Ibid., p. 19. [emphasis in original].

that candidates submit completed practice tests before being allowed to take the "Psychological Examination." In his "General Instructions to the Supervisor of the Scholastic Aptitude Test," he stressed that "A candidate is handicapped by not having had a practice booklet before taking the test, and it would be unfair of him to consider his score as comparable with scores of people who have had the practice booklet."<sup>11</sup>

Brigham was adamant about the importance of this requirement. For example, despite arrangements for all candidates to receive practice booklets a week before the examination, eleven candidates in 1926, eight boys and three girls, were admitted to the test without practice booklets. The psychologist stressed that these candidates had not taken the test under conditions comparable with those of the other candidates, and that he could not therefore report their scores. "It should be remembered that this action was made necessary because the scores obtained by candidates who had not studied the practice booklet could not be compared with the scores of candidates who had had ample opportunity to practice on the material of the examination."<sup>12</sup>

As revealed in their reports following the 1927 SAT, the test supervisors subsequently took Brigham quite seriously on the issue of required practice booklets. Extracts from Supervisor's reports in 1927 reveal this seriousness. One supervisor reported that a candidate identified as "R.E.M." applying to Yale University had not received his practice booklet because of a change in address;

---

<sup>11</sup> "General Instruction to the Supervisor of the Scholastic Aptitude Test," reproduced as "Appendix I" in Brigham, *A Study of Error*, p. 304.

<sup>12</sup> Brigham, *A Study of Error*, p. 331. [This is a significant requirement for Brigham to impose. He was indicating concern for what would later be called potential "error variance" by requiring preparation. Later E.T.S. would contend that no form of preparation or familiarity was necessary.

the supervisor reported that "it is not felt that he took the test under the same conditions as the other candidates."<sup>13</sup> Regarding a candidate "C.H.W." applying to Harvard, another test supervisor reported that although the lack of a practice examination meant that he had not taken the test under the same conditions, "nevertheless, he made a very high score." Brigham indicated to the supervisor that "this fact would seem to be worth reporting to the college for which he is a candidate, although the actual marks will be withheld as invalid in this case."<sup>14</sup>

The students' prior familiarity with the examination remained a concern with Brigham even after the Board itself became familiar with the new instruments. Attempting to streamline registration procedures in 1931, the Board modified its requirement that all candidates register in time to be mailed a practice examination. Brigham, however, was still not willing to allow candidates to take the SAT without a completed practice examination. To accommodate both administrative efficiency and Brigham's steadfast belief that the practice examination was crucial to accurate testing, the Board established compromise procedures. After 1931 the Board permitted test supervisors to register late applicants if they would still have at least fifteen hours prior to the SAT to complete the practice examination. "This procedure eliminated all those unpleasant situations in which the examinations of candidates had to be invalidated because of their having had no practice booklets."<sup>15</sup>

---

<sup>13</sup> "Extracts from Supervisor's Reports" Brigham Files, E.T.S. Archives.

<sup>14</sup> Ibid.

<sup>15</sup> "General Report on the Scholastic Aptitude Test," June 21, 1931, "Appendix E" in Thirty-first Report of the Secretary of the College Board, p. 186. The Commission clearly still saw the practice examination as critical. In its report, the committee felt it noteworthy to mention "that two candidates claimed to have lost their practice booklets between six o'clock Friday evening and nine o'clock Saturday morning, but as the records of the Board showed that these candidates

The requirement of a practice examination is significant in light of one of the longstanding controversies surrounding the Scholastic Aptitude Test. Educators and psychologists debate the question of the degree to which rehearsal and test familiarity can increase one's score. If the test measures something unchanging and is not influenced by culture and previous educational advantage, then candidates certainly could not prepare for the examination. Clearly, a test that is affected by short term rehearsal can not be measuring immutable or hereditary skills. By 1926 the Brigham, who had three years earlier overlooked completely the impact of a lack of test familiarity among immigrant soldiers taking the Army Alpha and Beta, was sophisticated enough to recognize the potential for "error variance" introduced by such test taking naivete.<sup>16</sup> His concern that candidates unfamiliar with the format of the items or of the test in general would be undermeasured indicated an awareness of limitations of the tests in consistently discovering talent.

#### The SAT As A Reflection of its Ancestry

Although the SAT was new to most members of the board and all of the candidates who first took it, it was not new in the context of the "psychological" examinations that the psychologists had introduced to higher education after the war. The test's choice of items, the format and the pacing, as well as the

---

had been mailed practice booklets, and as there was other evidence that they had studied their practice booklets, their examinations were scored and reported regularly."

<sup>16</sup> Prior to his report on the Army scores, Brigham had recognized one form of extreme error variance. In his doctoral dissertation on the Binet Tests, he had noted that American intelligence tests would not be appropriate for persons whose first language was not English. See: Carl Campbell Brigham "Diagnostic Value of Mental Tests," Psychological Monographs 24 (1917):131-133. An intellectual history of Carl Campbell Brigham is needed to pursue the subtleties of thought that could lead him from this recognition in 1917 to his 1923 work then back to a recognition of the impact of environment by 1926.

underlying psychometric assumptions, reflected the heritage of the Army Alpha and its descendents.

The most direct ancestor of the SAT was Brigham's own "Princeton Psychological Examination;" five of the nine "operating" sections on the first two SAT's came verbatim from the Princeton test.<sup>17</sup> Through this test, the SAT was directly connected to the Army Alpha examination. Brigham considered his Princeton test "largely experimental in nature."<sup>18</sup> In fact, however, this test was largely derivative, drawing heavily from the Army Alpha as well as other examinations of the early 1920s.

Though the Princeton examination's descent from the Alpha examination is clear, Brigham experimented, using novel item types and configurations; the examination was thus somewhat fluid. The innovative aspects of the Princeton examination came in Brigham's willingness to add sections and items eclectically to a general foundation established by the previous examinations. Brigham's use of the mechanical items involving visualization of gears and pulleys is an example of this experimentation. Inclusion of these mechanical items that measured attributes that would be difficult to classify as either verbal or quantitative distinguished Brigham's early work.<sup>19</sup>

The SAT was distinct from the Brigham's Princeton examination in its omission of items that did not fall neatly into either Verbal or Mathematical

---

<sup>17</sup> Brigham himself acknowledged the clear relationship between the tests. See: Thirty-Second Annual Report of the Secretary of the College Board (New York: College Board), p. 4. and; "The Princeton Test of 1925" in Brigham, A Study of Error, pp. 248-302.

<sup>18</sup> Brigham, A Study of Error, p. 248.

<sup>19</sup> This author is not here endorsing or criticizing the inclusion of such items. Their use is simply highlighted to show the early inchoate nature of the examinations.



categories. In that sense the SAT moved to a more orthodox position within psychometrics, by implicitly imposing a two-fold division upon the concept of intelligence. The first two SAT's were divided into two operating sections--verbal and mathematical--that together took ninety-seven minutes and an experimental section that took thirty minutes.

### The Verbal Section

Vocabulary is at the heart of the SAT verbal section; a strong vocabulary affects performance on each of the verbal item types.<sup>20</sup> The 1926 and 1927 Scholastic Aptitude Tests had seven sections that tested verbal skills. Five of these sections were more or less direct tests of vocabulary strength. These included sections called "definitions," "classification," "antonyms," and "analogies" and "paragraph reading." The remaining two verbal sections were "artificial language," "logical inference."

The first section facing students taking the 1926 and 1927 SAT's was one of the five copied directly from the 1925 Princeton Psychological Test.<sup>21</sup> The section presented thirty "definitions" in one column and then thirty words in another column and required the students, in a nine minute time period, to match the appropriate definition and word. For example, question three on both the SAT and on the earlier institutional test asked for the word defined by the phrase,

---

<sup>20</sup> The vocabulary dependent nature of the SAT is one of the bases for recent criticism. Cognitive psychologist Robert Sternberg notes that verbal intelligence tests are consistently based on strength of one's vocabulary. As a psychologist who is interested in the study of information processing, Sternberg wants to assess an individual's ability to learn words from context rather than simply how large that person's vocabulary. See: Robert J. Sternberg, "The Nature of Mental Abilities," American Psychologist 34 (March, 1979):214-230.

<sup>21</sup> This discussion of the use of specific item formats relies on information from P.G. Loret, "A History of the Content of the Scholastic Aptitude Test," Appendix VIII in 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.

"the projecting lower edges of a roof."<sup>22</sup> Brigham created these items based on Dartmouth psychologist H.T. Moore's work involving definitions.<sup>23</sup> Brigham, however, modified Moore's items making them "matching" exercises rather than open ended questions.

The second verbal section on the first two SATs consisted of twenty groups of six words. Charles L. Stone of Dartmouth had developed and standardized this "classification" test, which as presented on the SAT allowed eight minutes for twenty items.<sup>24</sup> This section too, had an unmistakable relationship to vocabulary. For each group the students were to choose the two related words. For example, a 1926 item gave students the words "silver, platinum, amethyst, coinage, emerald, and sapphire" and asked them to designate the two words that were "most closely related."<sup>25</sup>

The third verbal section on the first two SAT's required students to find fifty antonyms in a ten minute period.<sup>26</sup> Brigham himself developed this item type for the 1925 Princeton examination; he then modified it slightly for the SAT.<sup>27</sup> Antonym and Synonym questions have prima-facie the clearest relationship to vocabulary. No SAT has been offered that did not make use of at least one of these two most direct ways of testing vocabulary. The SAT item type in longest

---

<sup>22</sup> Loret, History of Content, Appendix B. p. 103, and Brigham, A Study of Error, p. 250.

<sup>23</sup> Brigham, A Study of Error, p. 248.

<sup>24</sup> Ibid.

<sup>25</sup> Loret, History of Content, Appendix B. p. 100. Loret does not reveal the answer to this question; the reader of his memorandum is left to muse over the relative "closeness" of the many possible relationships.

<sup>26</sup> Ibid., p. 26.

<sup>27</sup> Brigham, A Study of Error, p. 248

continuous use is the antonym which appeared in 1926 as a question in which students chose from among six options a pair of words that most closely represented the opposites, the item type was unchanged until 1952.<sup>28</sup> This original arrangement was a minor variation of an item configuration used in the Princeton examination; it became an accepted format. Brigham noted in A Study of Error, that the original form of this item was still (1932) being used in the American Council on Education Examination. The synonym was used in 1928 and 1929 then permanently dropped.

The fourth item type with an obvious relationship to vocabulary strength was the analogy item. Historically the analogy, like the antonym, has been a mainstay of the SAT. The English psychologist Cyril Burt first introduced the analogy item for use in intelligence tests in 1911.<sup>29</sup> Brigham modified the analogy item for the 1925 Princeton examination then placed it on his SAT in 1926.<sup>30</sup> With the exception of the years from 1930 to 1935, analogies have been in constant use. The analogy item has been presented on the SAT in three different forms. First, in 1926 and 1927, students were given the first three terms and asked to choose a fourth term that would produce a valid analogy. For example the candidate would be given "wire--electricity :: pipe-- (1 wipe, 2 tube, 3 water, 4 stem, 5 plumber)"; the task was to chose the answer that completed

---

<sup>28</sup> Loret, History of Content, p. 99. The 1952 change introduced the present format. A stem word is presented and a candidate finds the word that most closely approximates its opposite from among five choices.

<sup>29</sup> Cyril Burt, "Experimental Tests of Higher Mental Processes and Their Relation to General Intelligence." Journal of Experimental Pedagogy and Training College Record 1 (1911):93-112.

<sup>30</sup> Brigham, A Study of Error, p. 248.

the analogy.<sup>31</sup> In 1928, the SAT simultaneously used two forms of analogy items. Along with questions that presented three words, students faced questions that presented a pair of words and required them to choose two words from among six options that form a second pair with a relationship most closely analogous to the relationship. For example, "dawn : night -- (1.birth 2.cemetery 3.death 4.moon 5.life 6.sun)."<sup>32</sup> After a hiatus of six years during which the analogy was not used on the test, this second format was used through 1942. Beginning in 1943 and consistently to the present, the analogy question presented students the first ordered pair and asked them to supply a pair of words that had an analogous relationship. The consistent use of the analogy item reflects the psychologist's belief that these items measured an ability to reason.

As early as 1929 there were, however, those who feared the impact of coaching on analogy items. Hsuan Shan Chen, in a Teachers College dissertation found that analogy type items were "the most coachable test" among the verbal items.<sup>33</sup> Evidence on test coaching and preparation suggests that the more complex or novel the format of an item, the more coachable that item. If

---

<sup>31</sup> Loret, History of Content, Appendix B. p. 94. [Correct Answer is 3 (water)]. The 1926 examination also contained analogy items formatted differently in an experimental section. See Loret: p. 26. This specific item was, in fact, used earlier on the 1925 Princeton examination, see Carl Campbell Brigham, A Study of Error, p. 253.]

<sup>32</sup> Loret, History of Content, Appendix B. p. 92. [Correct answer is 1-3]

<sup>33</sup> Hsuan Shan Chen "The Comparative Coachability of Certain Types of Intelligence Tests." Ph.D. Dissertation Teachers College Columbia University, 1929. [See "Digests of Dissertations, September 1925-August 1929, Teachers College Library, p. 110.]

the difficulty of an item is in content and not in format, it is less coachable.<sup>34</sup> Brigham's early tinkering with the item format reflected an attempt to determine just what quality the item measured.<sup>35</sup> By 1960, E.T.S. was confident in its understanding of the responses of candidates to analogy items. Researcher Loret contends that the changes in formulation of analogy items eliminated "any puzzle element," allowing "the candidate to concentrate on meaning, and [not] requiring him to concern himself with coding the response or position."<sup>36</sup>

The one section of the first SAT that was written specifically for that test was called Paragraph Reading. This section also related to vocabulary in that it required candidates to find words that were incorrectly used. Based on earlier work by psychologist J.C. Chapman on tests to be used in early grades, Committee member Roswell P. Angier of Yale developed a form of reading comprehension test that required the candidates to read a paragraph and cross out words that kept the passage from making sense. The instructions on this section, used continually until major revisions of the test in 1941, stated that:

In each of the following paragraphs one important word, and one word only, has been substituted for another word and spoils the meaning of the paragraph. Find this word and cross it out: 1. A fondness for reading is an unfailing source of recreation. Whatever our dominant interest at the moment, we can find a book to

---

<sup>34</sup> See: Wild Cheryl L.; Spencer S. Swinton; and Madeline M. Wallmark, "Research Leading to the Revision of the Format of the Graduate Record Examinations Aptitude Test in October 1981." GRE Board Professional Report GREB No. 80-lbp, (November, 1982):1-20.

<sup>35</sup> See: Brigham, A Study of Error, pp. 65-162 for an example of his analysis of individual items--in the case he presented in those pages, it was analysis of mistakes with synonyms.

<sup>36</sup> Loret, History of Content, p. 101. For a different view of ETS success in perfecting the analogy item, see Owen, None of the Above, pp. 41-60.

harmonize with our mood, if we try; the friendliness of books is a fickle comfort to the book-lover.<sup>37</sup>

The two remaining verbal sections on the original SATs were far less directly related to vocabulary strength. However, neither of these two types of questions--"artificial language" questions, and "logical inference/premises and conclusions"--survived beyond the first two years of the test. The "artificial language" section, suggested by Stuart Dodd and developed by Brigham, asked the student to read a vocabulary and set of grammatical rules for a language.<sup>38</sup> The test provided the student with a group of contrived nouns and a group of verbs. The student was then asked to create sentences expressing particular meanings using such rules as "adverbs are formed by substituting the ending "id" for the "o" ending of the verb," or "future time is expressed by placing "bel" before the verb."<sup>39</sup>

David Camp Rogers provided the other shortlived verbal item type. Based on work he had introduced in 1919 for the Smith College Intelligence Test, "logical inference/premises & conclusions" items required candidates to read arguments while assuming "the premise or premises to be true and unquestioned," and then to judge presented conclusion. On the Smith College examination, the candidates simply chose among three possibilities; the

---

<sup>37</sup> Loret, History of Content, p.113.

<sup>38</sup> Brigham, A Study of Error, p. 248. See also Loret, History of Content, p. 9. Although this type of item vanished from the SAT, it has reappeared on other Educational Testing Service Tests. For example, during the 1960s and early 1970s, when the Law School Admission Test was still an ETS program, the LSAT occasionally included such items. Because this time period precedes Truth in Testing Legislation, this author is unable to ascertain whether these occurrences of these items were scored sections or experimental.

<sup>39</sup> Loret, History of Content, Appendix B. p. 98.

conclusion was "true, false, or undetermined."<sup>40</sup> For the SAT, Rogers made the answer choices more complex; the candidate had to indicate whether a conclusion was 1) necessarily true, 2) necessarily false, 3) probably true, 4) probably false, or 5) undetermined. A 1926 SAT question asked:

Premises:--The medals were given only to the best students, and Jacob received one of the medals.  
Conclusion:--Jacob is one of the best students.<sup>41</sup>

#### Mathematics: On, Off, Then On the SAT

The quantitative section of the SAT was initially problematic for the test's author. In selecting quantitatively based questions, Brigham sought to measure quantitative mental power, not simply computational ability; he purposely sought item types that did not reflect the content of any particular curriculum. Brigham initially was not certain whether mathematical items would reveal different traits than would verbal items. If intelligence was a single, unitary trait, quantitative sub-sections could be appropriately mixed in with verbal sub-sections to produce a single score. If intelligence was multi-faceted, as Brigham was increasingly predisposed to believe, the quantitative items should produce a separate score. The original 1926 examination included two types of quantitatively based items. The first quantitative section consisted of "number series" items that were a "new form of a widely used test developed and standardized by SAT committee member Charles L. Stone at Dartmouth."<sup>42</sup>

---

<sup>40</sup> David Camp Rogers, Smith College Intelligence Test, 1919, p. 2. Roger's Papers, Smith College Archives.

<sup>41</sup> Loret, History of Content, Appendix B. p. 108. Like to "Artificial Language" items that had a similar short tenure on the SAT, these items that Smith provided did not vanish entirely in 1927. A variation of these questions has appeared in the Analytical Ability Sections of the Graduate Record Examination and the Law School Admissions Test.

<sup>42</sup> Brigham, A Study of Error, p. 248.

The "number series" items appeared only in 1926 and 1927.<sup>43</sup> These items were distinctive to "intelligence tests," and quite new to non-psychologists. The questions consisted of a series of five numbers that reflected a series. The candidate had to produce the sixth and seventh numbers that would continue the same series relationship. For example, the 1926 test listed "4, 7, 18, 21, and 32" and asked what the next two numbers would be.<sup>44</sup> The novelty of its format made it the most coachable of the quantitative items.<sup>45</sup>

The first two SAT's also included a revision of the Princeton test's "Arithmetic Problems" section that Brigham wrote previously with Stuart .C. Dodd.<sup>46</sup> With some evident frustration, Brigham and Dodd had been collaborating since 1923 on an attempt to find quantitative items that would provide a "larger and fairer sampling of the subject's ability."<sup>47</sup> The "arithmetic problems" were primarily word problems in a free-response (non-multiple choice)

---

<sup>43</sup> Loret notes in his appendix to 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. 99, that the number series item reappeared in 1959 and 1960.

<sup>44</sup> Loret, History of Content, Appendix B. p. 110. Although Loret does not reveal the correct answer to this question, this researcher suggests that the last two numbers would be "35 and 46." The key to this question is to analyze the gaps between the numbers; they are 3 then 11. Practicing this skill, is in fact the basis of "coaching" such questions which are still used on the Doppelt Mathematical Reasoning Test published by the Psychological Corporation. For a rehearsal strategy see: Walter Pauk, How To Take Tests (New York: McGraw Hill, 1969), p. 79.

<sup>45</sup> Hsuan Shan Chen "The Comparative Coachability of Certain Types of Intelligence Tests." Ph.D. Dissertation Teachers College Columbia University, 1929. [See "Digests of Dissertations, September 1925-August 1929, Teachers College Library, p. 110.] Chen contrasts the number series item with items involving arithmetic. Arithmetic items, with their difficulty derived from content and not format were the least coachable items in Chen's study.

<sup>46</sup> Stuart C. Dodd to Brigham, April 27, 1923. Brigham Correspondence, E.T.S. Archives. See also: Brigham, A Study of Error, p. 248.

<sup>47</sup> Dodd to Brigham, April 27, 1923. Brigham Correspondence, E.T.S. Archives.



format. For example the 1926 examination contained the problem: "If a man's salary is \$20 a week, and he spends \$14 a week, how long will it take him to save \$300?"<sup>48</sup>

Brigham modified his views on what the tests accomplished based on evidence he gained in the earliest forms. The 1926 and 1927 tests had included both verbal and mathematical questions in a single score. After analysis of the 1927 test, the SAT commission concluded that the mathematical portion of the test should be discontinued. They allowed that it had "the germ of A.B. - B.S. differentiation and as a separate test will be of value to engineering and scientific schools," but they did not feel that it contributed to the usefulness of the SAT generally.<sup>49</sup> However, in 1929, Brigham reconsidered that conclusion and indicated that "the most important changes in the list of sub-tests were due chiefly to the discovery that the examinations of 1926 and 1927 included two distinct groups of sub-tests which, while of significance separately, might easily neutralize one another in combination."<sup>50</sup> By studying the intercorrelations among various sections of the new tests, Brigham reached the conclusion that the test might be measuring more than a single entity. The quantitatively based sections--arithmetical problems and number series--did not correlate strongly with the seven verbal sections. Cecyl Brolyer, who joined the Board's staff in 1927 and worked with Brigham as a statistician, later noted in an oral history interview that:

---

<sup>48</sup> Loret, History of Content, Appendix B. p. 97.

<sup>49</sup> Minutes of the College Entrance Examination Board, November 3, 1927, p. 7. College Board Archives.

<sup>50</sup> Twenty-Ninth Annual Report of the Secretary of the College Board, p. 22.

We came to the conclusion that we were losing information by adding the verbal and mathematical. You could have number X for individual A, that might represent high verbal, low on the number staff. For individual C, X might be the same score but result from low verbal and high mathematical. Now we wanted, when we discovered all this stuff, to separate them. We were trying to build two "pure" tests--high reliability, low correlation (i.e., between verbal and mathematical).<sup>51</sup>

In other words Brigham had discovered, by reanalyzing the earlier data, that the students obtaining good scores on the verbal questions might not be those obtaining good scores on the mathematical. The first response of the SAT commission was to drop the mathematical section altogether. Thus, for a two year period from 1928 to 1929, the SAT contained no specifically quantitative items.

In 1930, the SAT once again included a mathematical section. Although the SAT commission considered this reintroduction of the mathematical section merely experimental, it "hoped that this section would supply admission committees with supplementary evidence revealing whether the major interest of the candidate was mathematical or verbal, and that it might help to differentiate candidates into the 'liberal arts' and 'scientific' types."<sup>52</sup> Brigham noted that "in this way it will be possible to say of each candidate who takes the test that he has exhibited two distinct abilities or only one of these abilities or neither of them."<sup>53</sup>

---

<sup>51</sup> Gary D. Saretzky, E.T.S. Archives Oral History Program: Interview with Cecil R. Brolyer, June 13, 1984. Stuyvesant, New York. Educational Testing Services Archives, pp. 14-15

<sup>52</sup> Minutes of the College Entrance Examination Board October 29, 1930, p. 6. College Board Archives.

<sup>53</sup> Twenty-Ninth Annual Report of the Secretary of the College Board, p. 22.

The relationship of the abilities of mathematics and verbal skills continued to intrigue Brigham; he wanted to explore it further. Thus in 1929, along with proposing the reintroduction of the mathematical section as a separate entity, Brigham indicated that "A supplementary test of thinking in mathematical symbolism might serve the double function of revealing students who might profitably continue in the scientific fields, and of exposing those students whose proposed election of a scientific curriculum is the result of verbal disability rather than a high degree of mathematical and scientific aptitude."<sup>54</sup>

#### The Experimental Section: An SAT Innovation

A marked innovation on the SAT was the inclusion of an experimental section in which Brigham and his committee could evaluate new items. In fact, even the operating sections of the early SAT's administered between 1926 and 1930 reflected their experimental nature; item types were introduced, tried, modified and abandoned. In addition to these operating sections that themselves reflected the fluidity of the early examination, thirty minutes of each candidate's time was specifically devoted to completing an experimental section.<sup>55</sup> The testmakers did not identify the experimental section; they wanted candidates to complete the tasks with no knowledge that the particular items did not count. At the April 1, 1927 meeting of the College Board, Brigham reported that in the next year he would have 1500 items with which to experiment. He informed the Board that "A study of their answers enabled the Commission to detect certain types of

---

<sup>54</sup> Twenty-Ninth Annual Report of the Secretary of the College Board, p. 22.

<sup>55</sup> The inclusion of this section has formed the basis of some recent criticism. David Owen asserts that, because candidates are participating in experimentation, a human subjects release form should be required. Further, he criticizes candidate's having to pay to take a test then essentially working for the testing agency to develop future forms of the test. David Owen meeting with David Hubin, May 8, 1985.

response tending to lower the validity that could be discovered only experimentally."<sup>56</sup> During the first three years of his work with the SAT, Brigham experimented with "fifteen or twenty types of sub-tests, some of which were found to be unprofitable and were, therefore, discarded, whereas others were followed up and elaborated."<sup>57</sup>

Brigham's interest in research directly influenced the way he conceptualized and developed examinations. The pattern that Brigham established with experimental sections on the first SAT's Brigham's extended into all of his measurement work; his concept of test development involved experimentation and constant research. As Ellen Lagemann asserts, "few testers in the 1930s insisted upon constant experimentation to the degree that Carl Brigham did."<sup>58</sup> This was an expensive approach. The cost of preparing an experimental section and subtests for the examination "quickly runs into thousands of dollars."<sup>59</sup> The orthodox method of examination construction in the 1920s was to write items and then, based on the test developer's impressions or estimates, arrange them in an approximate order of increasing difficulty. Brigham led in the recognition that each item must have the power to differentiate among candidates. A useful item must be answered correctly by individuals who scored well on the test generally, and conversely missed by those who did poorly on the

---

<sup>56</sup> Minutes of the College Entrance Examination Board, April 1, 1927, p. 6. College Board Archives.

<sup>57</sup> Minutes of the College Entrance Examination Board, October 30, 1929, p. 6. College Board Archives.

<sup>58</sup> Ellen Lagemann, Private Power for the Public Good, p. 120.

<sup>59</sup> College Entrance Examination Board, Twenty-Sixth annual Report of the Secretary (September, 1926), p. 172.

examination. Brigham paid particular attention to wrong responses and pointed out that "for purposes of description, these wrong alternatives inserted to differentiate higher scoring people from lower scoring people may be called "traps."<sup>60</sup> Brigham paid uncommon attention to individual items. "The responses to each test-item included in the experimental sections of the scholastic aptitude test were plotted against the total score in the sum of the first nine tests in 1926 and 1927, and against the sum of the first seven tests in 1928."<sup>61</sup>

The range of item types included in the experimental sections of the early SAT's was extremely broad in terms of psychometric testing. Although the "operating" sections of the examination did not reflect Brigham's special interest in spatial and mechanical items, the experimental sections were definitely Brigham's. Because the experimental sections did not count toward a candidate's score, Brigham could be quite creative in what these sections contained. In some cases the early SAT experimental sections included items that reflected aptitudes other than verbal and quantitative. In other cases Brigham included sections of other examinations or even complete examinations. The first SAT contained an Otis Self-Administering Test, an aptitude test widely used to measure secondary school pupils.<sup>62</sup> After analyzing test scores, he concluded, however, that, in contrast to his test in general, the Otis did not differentiate among top candidates. He therefore dropped the Otis.<sup>63</sup>

---

<sup>60</sup> Brigham, *A Study of Error*, p. 367. Educational Testing Service no longer uses the term "traps;" the current term in use is "attractive distractor."

<sup>61</sup> Brigham, *A Study of Error*, p.366.

<sup>62</sup> Loret, *History of Content*, appendix A. p. 26.

<sup>63</sup> Carl Campbell Brigham, *A Study of Error*, p. 351. Brigham concluded that a perfect score on the Otis was equivalent to an SAT score of only 647.

Brigham had treated his Princeton test from 1923 through 1925 as an opportunity to experiment; he continued this attitude in his work with the SAT. The choice of items in the experimental section was often quite eclectic. Although a portion of the candidates received experimental sections that were indistinguishable from one of their operating sections, others received completely different types of items. The experimentation could range from relatively minor variations on question types used in the operating sections to items that appeared to be in no way related to those in the operating sections.

In 1926 and 1927 Brigham experimented with such items as "picture analogies" and "picture arrangement." In the "picture analogies" section the students saw a pair of drawings that expressed some relationship. They then had to choose two other drawings among six options that had "the same relationship."<sup>64</sup> "Picture arrangement" items required the candidates to infer a sequence among six pictures and then number the individual pictures.<sup>65</sup>

Administrative aspects of testing, however, constrained Brigham's experimental efforts. The testing schedule--a schedule that included the Board's comprehensive subject matter essay examinations--allocated only three and one-half hours for the SAT. The Board would not extend this so Brigham was forced to balance his interest in the mathematical section with his interest in experimentation with other items.<sup>66</sup> By 1930 Brigham and his associates in

---

<sup>64</sup> Loret. History of Content, Appendix B. p. 116.

<sup>65</sup> Ibid., p. 117.

<sup>66</sup> Minutes of the College Entrance Examination Board, November 2, 1932, p. 8. College Board Archives.

Princeton felt that they had an adequate backlog of tested items that they could drop the experimental section in favor of a mathematics section.<sup>67</sup>

For Brigham, dropping the experimental section in favor of the mathematical section was not desirable on a long term basis. He viewed the experimental section of the examination important to the long term development of the test. He wanted both a mathematics section and an experimental section. In 1932 Brigham requested approval of new time allocations. He wanted three hour time periods for "each division of the Scholastic Aptitude Test" to allow for the inclusion of experimental sections along with the Mathematics section. In his request for time changes, he noted that during the two year period in which there was no experimental section --1930-1932--"the material accumulated during the previous years had been used. This material is now exhausted and more must be collected."<sup>68</sup>

However, the leadership of the Board generally accepted the form introduced in 1930 and pressure developed within the board for the examination to become constant. In April of 1933, Board Chairman Radcliffe Heermance (of Princeton) appointed a commission, which he himself chaired, to consider Brigham's request for time periods to allow an experimental section.<sup>69</sup> A year later, President Marion Edgar Park of Bryn Mawr made a report for the "Committee on Examination Schedule." The committee had concluded after

---

<sup>67</sup> Brigham to James B. Conant, January 3, 1938. Archives, Educational Testing Service.

<sup>68</sup> Minutes of the College Entrance Examination Board, November 2, 1932, p. 8. College Board Archives.

<sup>69</sup> Minutes of the College Entrance Examination Board, April 5, 1933, p. 3. College Board Archives.

considering a "number of changes. . .that each of them was impossible" because of potential conflicts with the New York Regents examinations.<sup>70</sup> In 1936 Brigham succeeded in reintroducing an experimental section but only after protracted arguments with the Board that was seeking codification, and only by keeping the total testing time at the original three and one half hours.<sup>71</sup>

Although the first two SATs--those given in 1926 and 1927 were indistinguishable from each other in types of items and in format, Brigham's experimentation with the new test subsequently led to some early and rapid changes. The third version of the SAT, in fact, reflected general and significant changes. In addition to temporarily dropping the mathematical items, Brigham made changes in formats of questions within sections. For example he dropped the "Definitions" items used both in 1926 and 1928 in favor of what he called "double definitions." Mainstay sections of the first two examinations did not appear on the third test. The author dropped the "definitions," and "artificial language," "logical inference" sections from the "operating sections" of the SAT in 1928.<sup>72</sup>

The 1930 examination established--at least in the operating sections of the SAT--a predictable pattern; from that point forward, the SAT content that contributes to a candidate's score has been marked by very little change. In

---

<sup>70</sup> Minutes of the College Entrance Examination Board, April 11, 1934, p. 11. College Board Archives.

<sup>71</sup> Brigham to James B. Conant, January 3, 1938. Archives, Educational Testing Service.

<sup>72</sup> Although these items were quickly discarded from the SAT, they were not forgotten by the testmaker. For example, the "Artificial Language" item reappeared in the Law School Admissions Test in the early 1970's. The "logical inference" item reappeared with only minor alteration on the Law School Admission Test from June 1982 to the present.



contrast to early years of the SAT with its experimentation changes since 1930 have been less dramatic and generally merely evolutionary. For example in the thirties, after Warren G. Findley demonstrated in his doctoral dissertation that although results of all vocabulary tests are highly correlated, specialized content of the vocabulary words will influence test results, the vocabulary based sections of the SAT underwent a change of emphasis away from what ETS characterizes as "literary vocabulary" to vocabulary from a balance of disciplines.<sup>73</sup>

#### The Persistent Question of Speed and Power

Beginning with the early intelligence definition debates of the 1920s, and continuing to the present, psychometrists and the lay public have differed over the role that a person's speed at completing tasks should be reflected in a test score. E. G. Boring had justified "speeded" tests because they would measure mental horsepower.<sup>74</sup> Others, ranging from Walter Lippmann and Leon Thurstone in the 1920s to Robert Sternberg in the 1980s have criticized the use of speed as a differentiating factor. Sternberg has argued that the current tests, with their psychometric heritage from the Army Alpha, proceed on the assumption the "smart is fast."<sup>75</sup> Speededness as a factor in intelligence testing

---

<sup>73</sup> Warren G. Findley. Specializing of Verbal Facility at the College entrance Level New York: Teachers College Contribution to Education No. 567, pp. 52-53; Coffman, William E. "Research and Development Report--The Scholastic Aptitude Test - 1926-1962" Test Development Report TDR-63-2, June 1963. E.T.S. Archives, p.11

<sup>74</sup> Edwin G. Boring, "Intelligence as the Tests Test It" New Republic, (June 1923):36. Historically, a discipline that accepted Boring's metaphor of the mind as an engine--power of the mind was indicated by the ability to work in a 'high gear' without downshifting--was not influenced by the definitions of intelligence that valued thoughtful reflection.

<sup>75</sup> Robert J. Sternberg quoted in Goleman, Daniel. "Rethinking the Value of Intelligence Tests" The New York Times Education Life, (November 9, 1986):27. Sternberg, as a psychologist critical of I.Q. tests notes, that "what is critical. . . is not speed per Se, but knowing what to do quickly and when to act slowly. . .snap judgments are often poor judgments.

came under criticism as early as 1928, when Otto Klineberg, an anthropology student of Franz Boaz reported on his observations of Yakima Indian children: "The Indian children appeared almost completely indifferent to the amount of time required to complete the tasks. No matter how often I repeated or emphasized the words 'as quickly as possible' they paid no attention. On the other hand, they made fewer mistakes."<sup>76</sup> As Raymond Fancher points out, Klineberg had found a culture which "the most intelligence people, by their own standards, would be actually *penalized* on all intelligence tests items requiring speedy responses."<sup>77</sup>

Writing in 1926, Carl Brigham explained his perspective on the importance of speed on a test.

It is particularly difficult for the person taking a group test to understand why he is not allowed to finish any test he has started. Time limits of single tests are usually fixed so that the average of the whole group will be just about one-half of the total number of problems. In other words, if a test contains forty problems, the time limit is set so that the average of all persons taking the test will be approximately twenty questions right. The purpose of fixing the time limits in this way is to give the best men room to display their ability.<sup>78</sup>

Brigham viewed speed as important but recognized that earlier tests had stressed speed to the exclusion of other attributes. Brigham pointed out that in the easy examinations like the Alpha, the only differentiating factor on certain

---

<sup>76</sup> From Otto Klineberg, Autobiography, in G. Linzey, ed., A History of Psychology in Autobiography 6 (1974):167. (Englewood Cliffs, N.J.: Prentice Hall,. quoted in Raymond Fancher, The Intelligence Men: Makers of the I.Q. Controversy, New York: W.W. Norton, 1985, p. 131.

<sup>77</sup> Raymond Fancher, The Intelligence Men: Makers of the I.Q. Controversy, New York: W.W. Norton, 1985. Fancher presents this in the context of his discussion of the work of Klineberg and Boas in presenting the 'Environmentalists' Response' to the Hereditarian thesis. see pp. 130-132. [Emphasis in original].

<sup>78</sup> Carl C. Brigham. "Intelligence Tests: The Third of the Present Series of Princeton Lectures by Members of the University Faculty" Princeton Alumni Weekly 26 (May 5, 1926):788.

sections was speed. "Revisions of tests for college purposes stress more the difficulty or depth of the items, so that more liberal time may be allowed. Ten, fifteen, and twenty minute time limits may be used in college tests rather than the three and five minute tests of army alpha. Distance travelled into the test is more of a measure of insight than of speed."<sup>79</sup>

Brigham sought a test that would measure verbal "power" as opposed to speed. But, despite Brigham's criticisms of the Alpha for their excessive speededness, the early SAT's were highly "speeded" tests. The SAT that the students first took in 1926 took three hours and consisted of ten sections containing 380 questions.<sup>80</sup> In the analogies sections, candidates had six minutes to answer forty questions. In his 1927 report on the test, Brigham indicated that a test that was highly speeded did not appear to measure the same qualities as a test that allowed candidates more time. Therefore, in the third form of the test, Brigham provided "much more liberal time limits."<sup>81</sup>

In his 1961 study of SAT content, Loret reported several "general trends in the development of the SAT."<sup>82</sup> One of the trends was the gradual shift away from "highly speeded" items. The changes indicated have been of a relatively

---

<sup>79</sup> Ibid.

<sup>80</sup> In contrast By 1958, the test still took three hours, but the number of sections had been reduced to five and the total number of questions had been reduced to 150. [It should be noted that even that latter examination is considered a "speeded test" with an underlying assumption that only one-half of any given group of candidates will finish a given section.] Peter G. Loret, "Test Development Memorandum: A History of the Content of the Scholastic Aptitude Test" ETS Archives TDM-60-1, October 1960, p. 80.

<sup>81</sup> Carl Campbell Brigham, A Study of Error, p. 351.

<sup>82</sup> P.G. Loret in 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. 102

gradual nature but the growing emphasis on providing the candidate sufficient time to complete the test is nevertheless quite obvious.<sup>83</sup>

Trends in timing, however, are less influenced by theoretical changes in the discipline of cognitive psychology or in changes in learning theory than they are by practical considerations.<sup>84</sup> The SAT as a selection instrument, not a diagnostic instrument must rank candidates and arrange them in a statistically predictable pattern. In developing an instrument that will report the anticipated "normal curve" of subject performance, one of the easiest ways to assure a "proper" distribution is to make speed a factor. Thus, if the intent of the SAT is to distribute candidates along a curve to allow admissions officers to justify their decisions, it is of little consequence how that curve is derived.<sup>85</sup> As long as the basis for the curve is no obviously antithetical to the quality that the test is meant to predict, there is no marketing imperative for change. In fact, marketing forces could counter impulses to change.

---

<sup>83</sup> Ibid., p. 103. See also 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.9. The pattern of reducing speededness, however was subsequently reversed in 1972 with the introduction of a more speeded version of mathematics in order to allow time for the new Test of Standard Written English. ETS still considers the SAT to be a relatively unspeeded test. See: "The Effects of Coaching on Standardized Admissions," Staff Memorandum of the Boston Regional Office of the Federal Trade Commission (September 1978), p. 1.

<sup>84</sup> Educational Testing Service replaced, in one section of math items, a traditional five-option "discrete quantitative" in which candidates chose a value that answered a question, with "Quantitative Comparison" items that asked for comparisons of values. This was necessary to allow time for students to take the Test of Standard Written English in the same testing period as the SAT.

<sup>85</sup> A simple example of the effect of speed on producing the bell curve sought for admissions purposes would be to ask students to complete a series of one-hundred two digit addition problems. If enough time were allowed most candidates could complete the task with only trivial errors. If, however, a time limit of three minutes were imposed, a bell curve of performance would emerge.

Just as the ranking purposes of the SAT influenced its construction regarding speed, so too did they effect the order of items. Consistent with generally accepted theory underlying intelligence and aptitude testing the Scholastic Aptitude Test presented items in order of increasing difficulty. Based on the early work of Binet who used ascending order of difficulty to indicate the point at which a student could no longer provide accurate answers, intelligence testers have consistently arranged items with the easiest items first.<sup>86</sup> Generally the arrangement reflected the test designer's prediction of the actual difficulty of items.

For the SAT, Brigham and his assistant Brolyer developed a different approach to ascertaining the difficulty of items prior to arranging them. Based on their experimental sections and experimental items within operating sections, Brolyer and Brigham assigned what they called a "delta" value to each question after they saw the way people responded to it.<sup>87</sup> Then, in developing subsequent examinations, they would draw from their pool of questions based on a particular question's delta value. If experimentation with a section indicated that it was too difficult, they would replace several high deltas questions with easier questions.<sup>88</sup>

---

<sup>86</sup> For a discussion of this aspect of "modern" tests, see Joseph D. Matarazzo, Wechsler's Measurement and Appraisal of Adult Intelligence, (New York: Oxford University Press, 1972) pp. 250-252.

<sup>87</sup> "Cecil Brolyer, Right-Hand Man to Carl Brigham Between 1927 and 1936, remembers SAT's Infancy," ETS Examiner 14 (February 21, 1985):1.

<sup>88</sup> An item is considered difficult on the SAT if a large percentage of people miss it. The content of that item may not, in fact, be particularly difficult; the testtaker's missteps may result from one of the answer choices being a particularly "attractive distractor."

### Test Security

As early as June of 1925, Brigham made clear his plans not to reveal the test either to the public or to other psychologists. He indicated that "the examinations to be prepared by the College Board must be held secret until each issue is dead, and therefore can not be made available to members of the Psychological Association generally until after the examinations have been given."<sup>89</sup> Brigham justified the security of the exam in part on financial ground. His 1926 report stressed the cost of producing the examination but said that "After the usefulness of this particular examination has been exhausted, it will be released."<sup>90</sup> Because Brigham ended up braiding items from one test into the next, the test security that Brigham proposed essentially meant that the SAT was not available at all for scrutiny by other psychologists.

Keeping the test secure was no trivial matter to the SAT's author. Brigham considered the issue significant enough to consistently devote space to it in his annual reports to the Board--reports that covered general philosophical and administrative issues in the administration of the test. In 1927, he was concerned that "at the time of writing this report, the check on missing booklets has not yet been completed, but the situation this year is undoubtedly worse than it was last year. At least seven booklets were lost without any chance of recovery."<sup>91</sup> His

---

<sup>89</sup> Brigham to Yerkes, June 16, 1925, Yale University Archives, Robert M. Yerkes Papers.

<sup>90</sup> Twenty-Sixth annual Report of the Secretary of the College Board, (September, 1926):172. See also: Minutes of the College Entrance Examination Board, November 3, 1927, p. 8. College Board Archives.

<sup>91</sup> Brigham, "General Report on the Scholastic Aptitude Test, June 25, 1927." Twenty-Seventh Annual Report of the Secretary of the College Board, p. 188.

1928 report showed that security had improved and that only "two books were lost without any chance of recovery."<sup>92</sup>

That the SAT was not revealed to other psychologists created tension and mistrust. Speaking in 1985, Ben D. Wood attributed his not seeing an SAT to his close relationship to Thorndike and his opposition to Brigham on professional issues. Downplaying the need for test security as "the excuse they had," He noted:

I suppose the inevitable results of the disputations was that I was never permitted to see any single SAT. I've never seen one. And that could not be just accident, that was deliberate withholding. I don't know why. We published our objective tests and anybody could look at them and criticize them. . . The attitude of Brigham and his group was never to show us anything that was being even contemplated by them for their present or their current or the future SAT's.<sup>93</sup>

In 1938 Walter Van Dyke Bingham wrote to Brigham criticizing the policy of not making the SAT available to other psychologists for research. "What have you decided to do about making available for general use in secondary schools one or more forms of the Scholastic Aptitude Tests.? You have developed the best examination there is for measuring academic aptitude. That the measurements which such an instruments yields are not readily accessible to high school principles and other advisors when counseling students about their educational plans is in my opinion one of the scandals of the present situation in the testing movement. I wonder whether the College Entrance Examination

---

<sup>92</sup> Brigham, "General Report on the Scholastic Aptitude Test, Twenty-Eighth Annual Report of the Secretary of the College Board, p. 161.

<sup>93</sup> David R.Hubin, "Oral History Interview with Ben D. Wood." Croton on Hudson New York, May 23, 1985. Tape and Transcript on file at E.T.S. Archives, p.9.

Board realizes that in not releasing such examinations it has in part been responsible for the continued wide use of inferior instruments."<sup>94</sup>

In 1941 the Board did, for the first time, approve a request of another psychologist to make use of the SAT. Robert L. Thorndike of Columbia's Teacher's College requested permission to use the Scholastic Aptitude Test "in a study of intelligence" which he was planning to make. The Board approved the request with the condition that the tabulation of scores be made by an employee of the Board and that; information "in regard to individuals, schools, or colleges be considered confidential;" and that "any report based on the proposed investigation be filed with the Board in advance of publication."<sup>95</sup>

#### The Scoring Processes and Scales

The first SAT's, and in fact all SAT's through the 1940 testing year provided scores that were not equated from one year to another. Each year the mean score of the participating students was set at 500 on a scale from 200 to 800, and the standard deviation was set at 100. From 1926 to 1929 the Board reported only a single score; from 1930 through 1935 the Board reported separate verbal and math scores on the 200 to 800 scale. Score reporting between 1935 and 1941 varied as test configuration changed.<sup>96</sup>

Treating each testing period as a separate unique episode provided the freedom that Brigham enjoyed to experiment with different forms, items, and pacing. This was a freedom that would vanish after the 1941 codification of the

---

<sup>94</sup> Quoted by Brigham in a letter to James B. Conant January 3, 1938. E.T.S. Archives. Brigham File.

<sup>95</sup> Minutes of the College Entrance Examination Board, April 2, 1941, pp. 6-7 College Board Archives.

<sup>96</sup> Loret, History of Content, p. 9.



SAT as the College Board implemented a complicated system of equating scores from one year to the next..

The scoring process itself was labor intensive, preceding the invention of machine scoring; many of the items were free-response. Therefore Brigham and his assistant Brolyer hired a group of students who worked out of offices at the edge of the Princeton campus. Brigham reported:

The physical conditions for scoring were ideal. The clerks used large drawing-rooms which were adequately lighted and ventilated. Other rooms were available for receiving records, checking, sorting, and other work, and were well adapted to the various purposes. The Sundstrand Adding Machine Company loaned the scoring unit five adding machines. Two Monroe Calculators were available."<sup>97</sup>

The SAT scale of 200 to 800 became dominant in College Board administered tests.<sup>98</sup> In December, 1947 the College Board decided not to report scores that fell below 200 or above 800. The scores of the very top candidates have always been a concern to the College Board. "Students at this level of ability are always the object of special notice, particularly so in these days when unusual interest is being focused on high-level intellectual talent."<sup>99</sup> As E.T.S. research William H. Angoff noted in 1961, that this is a one-sided restriction, that is, that it does not ensure that there will always be 800-scores (or

---

<sup>97</sup> Brigham, A Study of Error, Page 332.

<sup>98</sup> This scoring scale, with its apparent precision has been the focus of criticism. David Owen, notes that colleges infer actual differences between candidates with inconsequential score differences. Owen recommends that the SAT be scored on a five point scale. See: Owen, None of the Above, p. 282.

<sup>99</sup> William H. Angoff "Implications of the 800-Score Problem for College Board Scaling and Equating" August, 1959 Memorandum. Appendix XIII. in 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. 178.

200-scores), permits considerable variation from one test form to the next."<sup>100</sup> Since variation will, in Angoff's words, "occur even if groups taking the different forms are identical in ability," scores would appear less stable.<sup>101</sup> Thus, reporting scores above 800 would highlight variability in the test.<sup>102</sup>

Initially the Board reported SAT scores only to the colleges to which candidates indicated they were applying; neither the students themselves, nor their secondary school instructors were given any information on their performance. In 1937, the Executive Committee of the Board voted to send the scores of candidates taking the SAT early to their schools because the scores might be of some use in advising students on colleges.<sup>103</sup> Making public to the schools the scores of groups of candidates established the foundation for competition among schools, among districts, and even among states.<sup>104</sup>

#### Member Institutions React to the SAT

The institutions that the students chose for score reports reflected generally the membership of the Board. However, not all institutions were ready to require or even request the submission of a score. Records indicate that the

---

<sup>100</sup> *ibid.*, p. 180.

<sup>101</sup> William Angoff, in Wilks, Scaling. The quoted part of the sentence is Angoff.

<sup>102</sup> Reporting scores in this highly visible group in which there was bound to be fluctuation would give the appearance of instability in the test. Or, alternatively, if educators had complete confidence in that the test as a measure never changed, such fluctuation at the top would cause some to infer a fluctuation in the aptitude of the American people.

<sup>103</sup> Minutes of the College Entrance Examinations Board, October 27, 1937. College Board Archives. Not until 1956 did the Board make scores of the SAT available to candidates themselves. For a discussion of this change, see: William G. Fels. "Who Should Know the Score?" College Board Review 30 (Fall 1956):16-17.

<sup>104</sup> As with many aspects of controversies surrounding the SAT, score reporting has ironic aspects. Reporting scores to students and schools was a reform, yet it contained the seeds of the invidious comparisons that many educators would consider an abuse of the test.

candidates taking the new test were applying to the exclusive private institutions-- primarily Ivy League Schools. It is noteworthy, however, that six candidates in 1926 took the examination for the University of Michigan; one for Indiana University, and one for the University of Wisconsin.<sup>105</sup> The 1927 distribution of candidates was similar to that of 1926. However, among those institutions that sent one candidate there was a broader geographic distribution; those institutions included Cal Tech, The University of Southern California, Purdue, and the University of New Mexico.<sup>106</sup>

Although most members of the College Board reacted favorably, or at least indifferently, to the use of the Scholastic Aptitude Test in 1926, there were two significant exceptions. Initially Harvard and Bryn Mawr refused to participate in the new SAT "on the grounds that they are unintelligent rather than intelligent."<sup>107</sup> Harvard initially led the opposition to the SAT. John M. Stalnaker, who joined the Board as a psychologist in 1934 and later became president of the National Merit Scholarship Corporation, noted that at Harvard, "the opposition of a strongly entrenched Old Guard was very great."<sup>108</sup> In the 1926 meeting of

---

<sup>105</sup> Brigham, "First Annual Report of the Commission on Scholastic Aptitude Tests" Twenty-Sixth Annual Report of the Secretary, p. 162. [Brigham notes that his data are compiled from the statements of the candidates on their test booklets.]

<sup>106</sup> Brigham, "Second Annual Report of the Commission on Scholastic Aptitude Tests" Twenty-Seventh Annual Report of the Secretary.

<sup>107</sup> Philadelphia Public Ledger, November 12, 1925. See also: "Intelligence Tests Barred by Harvard: Are Held Too Unreliable" New York Evening Post November 11, 1925. Clipping in College Board Archives.

<sup>108</sup> John M. Stalnaker to Mrs. Ralph J. Sharp, Brigham Papers. ETS Archives. [Stalnaker, president of National Merit Scholarship Corporation, writing with comments on Downey biography.]

test supervisors Thomas Fiske, after predicting great success for the exam, noted that

A month ago, Mr. Pennypacker, director of the Harvard Committee on Admission, wrote that Harvard was not interested in the psychological examination, and that no candidate for admission to Harvard would take the psychological examination. Yesterday I got a letter from Mr. Pennypacker saying that the Harvard Admissions Committee had voted to request all candidates for admission to Harvard to take the psychological examination.<sup>109</sup>

This was not to be, however. The Harvard faculty rejected the idea of requiring the new Scholastic Aptitude Test. Secretary Fiske inferred that the Harvard committee simply was not making the new examination a requirement but that it would use the information if provided. Fiske reasoned that certainly "if a boy takes it, the Harvard Committee on Admissions will try to use the additional evidence which is furnished by the psychological examination, and so they will advise boys who are coming to Harvard that they had better take it. Every little [bit] helps in the way of information."<sup>110</sup> James B. Conant comments, "It was hardly a new invention when we decided at Harvard to employ it as an instrument to differentiate among the incoming freshmen. Yet the more I learned about the use of the new objective tests and the more I became familiar with the concept of scholastic aptitude, the more I showed signs characteristic of a recent convert to

---

<sup>109</sup> Minutes of the Conference of Supervisors Held in the Trustee's Room, Columbia Library, Saturday, January 9, 1926, p. 8. Educational Testing Services Archives.

<sup>110</sup> Ibid.

a new religion.<sup>111</sup> Conant recalls that at the time, he felt that the SAT was "probably the best of several tests that were in use at that time."<sup>112</sup>

Although Harvard did not participate in requiring the SAT, many of its applicants were, in fact, simultaneously applying to schools that did require the new test; Columbia, Yale and Princeton made early use of it, so "Harvard had a good many students who took it either because they just wanted to take it or because they were also applying to other institutions that required it so that Harvard did get a number of scores reported."<sup>113</sup> Chauncey reports that "there had been enough students taking it so that I could make some studies to see how well the SAT predicted."<sup>114</sup> Significant numbers of Harvard students thus took the SAT but the institution did not require the test of all students until 1933.<sup>115</sup>

Aside from Harvard and Bryn Mawr, colleges generally embraced the new test, albeit in some cases with some misgivings. Dean C. Mildred Thompson of Radcliffe was more impressed by the SAT than were her Harvard counterparts but she too had was not without some reservations. After receiving her first score report from Brigham she responded by explaining how the SAT was used.

---

<sup>111</sup> Conant, James B. *My Several Lives: Memoirs of a Social Inventor*, New York: Harper and Row, 1970, p. 418 From the vantage point of his 1970 autobiography, Conant comments, "I was too ready to see in two new concepts--verbal aptitude and mathematical aptitude--the keys which would unlock all doors to a more promising future. (p. 419).

<sup>112</sup> Ibid.

<sup>113</sup> Gary D. Saretzky with Henry Chauncey, "Oral History with Henry Chauncey" March 28, 1977, p. I-10 and I-11. E.T.S. Archives.

<sup>114</sup> Ibid., p. I-10.

<sup>115</sup> Ibid.

In regard to the S.A.T. records we have used them primarily for guidance at the two extremes. If a student had a very high S.A.T. rating we considered very seriously before we eliminated the student from our entrance list, and in border line cases a low S.A.T. rating would be sufficient to throw the student out of the accepted list. In the great bulk of perfectly competent minds I think it extremely difficult to find a test which can differentiate grades of ability. If you can devise such it will be of great help to us. Our task of selection becomes more serious each year and when we have at least a hundred candidates on the border line who are actually ready for college as far as test [sic] can indicate, any additional inquiry to throw light upon the capacity more than training would be of the greatest assistance to us."<sup>116</sup>

Thus for Thompson, the SAT while of some use, had not solved the problem of discerning the future abilities of students to learn.

#### Press Reaction to SAT

The nation's press generally paid little attention to the introduction of the Scholastic Aptitude Test. It did, however, comment on the Harvard's reaction. That University's refusal to require the new SAT as a part of its admissions process was received well by some members of the press; The New York Sun in an Editorial on November 13, 1925, commented: "It is reassuring and encouraging to have Harvard University and Bryn Mawr College decline to follow their associates. . . in adoption of psychological tests to establish the desirability of admitting applicants for instruction. . . The disciples of the great God of Standardization have adopted psychology in all its manifestations as their own and it takes a bold man or woman to oppose them."<sup>117</sup> The New York Times, similarly opined that "It was probably inevitable that Harvard and Bryn Mawr

---

<sup>116</sup> C. Mildred Thompson to Carl Campbell Brigham, September 3, 1927. Archives, Educational Testing Services. [Filed with Early Score Reports].

<sup>117</sup> "A Flash of Independence" New York Sun, November 13, 1925. Editorial. [Clipping on File at College Board Archives.]

would be the two educational institutions that would first raise a standard of revolt against the cheap chatter of modern psychology."<sup>118</sup>

In a few cases, the press not only discussed and evaluated Harvard's rejection of the SAT, it offered its own judgments of the usefulness of the examination. The Philadelphia Bulletin, after reporting that the College Board was implementing the new tests, cautioned that "the tests may indicate the speed with which, on a given occasion, subjects react to suggestion: but what they show about fundamental intellectual quantity is yet to be determined."<sup>119</sup> These observations from the press notwithstanding, the first administration of the SAT must be considered a success; the College Board and the majority of the admissions officers associated with the board, considered it so.

### Conclusion

The Scholastic Aptitude Test that Brigham and his committee introduced in 1926 was a derivative instrument drawing heavily from existing examinations-- particularly the Princeton Psychological Examination. The first years of the SAT's existence involved extensive experimentation with test format, item choice and pacing. Part of the freedom to experiment was created by the very fact that the SAT occupied a clearly secondary role in College Board testing.

Entering the 1930s Brigham's latitude to experiment progressively diminished as the test became codified by its own success. In the next chapter we will examine the forces that led to codification that resulted ultimately in a

---

<sup>118</sup> "Two Sensible Colleges." New York Times: November 14, 1925, See also: "Colleges To Try 'Intelligence Tests,'" The New York Times, November 11, 1925. and; "Ban 'Unintelligent' Tests." The Philadelphia Ledger, November 12, 1925. [Clippings on file at College Board Archives.]

<sup>119</sup> "Intelligence Speed Tests" The Philadelphia Bulletin. November 14, 1925. [Clipping on file at College Board Archives.]

highly predictable, ossified instrument after 1941. The advent of machine scoring contributed to codification. So too, did the increased pressure on admissions officers resulting from demographic changes in higher education after World War II.

Suppleness would vanish in the hands of the marketers who faced admissions officers hungry for established norms and absolute statements. Ironically, the marketing success of the SAT would provided Brigham with the wherewithal to experiment while simultaneously circumscribing the effect of his experimentation. Thus from the perspective of Carl Campbell Brigham, the experimenter, the first decade of the SAT's existence had given him the freedom to develop an instrument that he hoped to keep supple; this was a freedom that Brigham thought essential but his battle to preserve that freedom would be a losing one.