CHAPTER VII

THE CONFLICT: WIDE USE OR WISE USE OF TESTING Introduction

The College Board, along with the country at large, entered a troubled decade as the 1930s began. For the Board, the thirties were a general period of turmoil as it adjusted to new roles; the number of candidates taking its traditional examinations fell as the national economic depression made it impossible for many students to consider college. The decrease threatened to relegate the Board to regional influence among a small number of elite schools. The country's appetite for tests, however, was not diminished; educators called for new tests and new uses for existing tests. Faced with a market for new tests, the Board and other test developers had to consider the relationship of research to marketing. During the decade of the 1930s, Brigham, viewing the tests as experiments for careful, limited use, became an outspoken critic of "wide use" of the tests.1

By the mid-thirties the distribution of tests had reached a level that prompted the American Council on Education to establish a "Committee on Review of the Testing Movement." In these discussions, Brigham took the position that testing theory and research should precede wide distribution of tests; others, most notably Ben Wood, felt that wide distribution would provide

¹ I am indebted to Robert Hampel of the University of Delaware for the phrase "wide use versus wise use" as used in the title of the chapter. Hampel used the phrase in answer to a question after an oral presentation "Institution, The Origins of ETS," at Educational Testing Service, Princeton, New Jersey, June 19, 1988.

the data for further improvement of the tests. In their own ways, both groups were seeking progress in testing but both viewed the opposing positions as directly antithetical to progress. As the decade of the thirties progressed, Brigham lost ground to his opponents.

The growth of the Scholastic Aptitude Test in the thirties, although quite modest by standards of the next decade, contrasted sharply with a constant decline in applicants for the traditional Board essay examinations. Board members saw the SAT's growth as evidence of its success and increasingly equated such success with the Board's very existence. The Board's perception decreased Brigham's latitude to experiment with the instrument. Brigham did enjoy a victory at the end of the decade as he fended off a proposal to form a national testing agency, but ultimately, in the next decade, the promoters of broad extensions of testing would not be denied.

Psychology and Testing in the 1930s

By the late 1920s and early 1930s, attitudes among psychologists were shifting away from a virtually unanimous acceptance of the hereditarian concept of intelligence; psychologists were increasingly at least entertaining consideration of the role of environment in shaping intellect.² Further, in the early thirties a more mature discipline of psychology abandoned some of the stridency that had characterized the debates of the previous decade, and fine points of theory could increasingly be debated without acrimony.³

² See for example Arnold Gesell's testimony to the National Research Council, 1933: Transcript in Arnold Gesell Papers, Box 142, Library of Congress.

³. For a discussion of the dialogues of this period, see Britell, <u>Never Quite a Public Dialogue</u>, pp. 231-234. Britell characterizes this period as one in which a "dialogue" on testing almost developed.

Thus psychology, although disparate in its components, was not clearly divided; there was one general discipline of psychology. Warren G. Findley asserts that during this period "one could be a psychologist, pure and simple."⁴ Although distinctions of emphasis existed within the discipline, no officially acknowledged divisions existed between theoretical and applied psychologists.⁵

Also in the early 1930s psychologists embarked on a new period of extensive psychometric research as they explored new theories of intellect and the relationship of special abilities to general ability. Simultaneously, however, the marketing processes that would lead to "wide distribution" of the tests set in motion an ossifying process that separated the tests from advances in the field of psychology. By the end of the decade of the thirties and then in the immediate aftermath of World War II, the marketing aspects of the testing industry were so overwhelming that there was little likelihood that an instrument like the SAT could develop as a supple instrument that could reflect advances in the disciplines of education or psychology.

Brigham as Associate Secretary

Brigham, himself, emerged from the 1920s as one of the leading researchers in concepts of aptitude testing. In 1930 Brigham persuaded the College Board to form a Research and Statistical Laboratory in Princeton and, in

⁴ Warren G. Findley, "Carl C. Brigham Revisited" <u>The College Board Review</u> 119 (Spring 1981):9.

⁵ Divisions would develop in 1938 when applied psychologists formed the American Association for Applied Psychology. The AAAP later merged with the APA as a branch. For a discussion of the potential division among psychologists, see: Franz Samelson "The 'Impending Dismemberment of Psychology' and its miraculous rescue, 1930-45." Paper presented at 20th Annual Meeting of Cheiron, Princeton, New Jersey, June 20, 1988.

⁶ Britell, <u>Never Quite a Public Dialogue</u>, p. 253. Guilford and Thurstone led in research on this question.

April of that year, the Board appointed Brigham as the full-time Associate Secretary for Research.⁷ He took a leave from his faculty position at Princeton to work full-time on measurement research for the Board. With the position finally approved in October of that year, Brigham became a full-time staff member of the Board as Associate Secretary. Brigham saw his work with the Board as a chance to do research on the development of testing. As the Board finalized arrangements with Princeton University for Brigham to take a leave, the psychologist wrote to his friend Yerkes indicating that "Princeton is cooperating beautifully in the venture and I am just about to create my own little Institute for Educational Research, although it may not be called that."

Two themes emerged in Brigham's work as Associate Secretary for the Board. First, he advocated strongly that an experimental and flexible approach characterize Board activities: "The Board assumes a task that will never be completed. It takes on the character of a function and not a constant." Second, Brigham sought to experiment with developing tests for purposes other than admissions; his early interest in the Princeton Bogey concept continued and developed into an interest in examinations for placement purposes rather than simply for selection purposes. With implications for a potential national testing agency that he would later vehemently oppose, he observed that "the Board

⁷ Fuess, <u>The College Board: Its First Fifty Years</u> (New York: College Board, 1967), p. 110.

⁸ Carl Campbell Brigham to Robert M. Yerkes April 16, 1930. Robert M. Yerkes Papers, Yale University Archives. [Brigham here is certainly being humorous about the name, given that the title would have duplicated Thorndike's at Columbia.]

⁹ "Report of the Associate Secretary" in <u>Thirty-Third Annual Report of the Secretary of</u> the College Board (New York: College Board, 1933), p. 14.

might continue as a College Examination Board, but not necessarily as a College Entrance Examination Board."¹⁰

At the end of 1932, the Board asked Brigham to explore ways that Board examinations could be more useful for placement purposes. In his 1933 report to the Board as Associate Secretary, Brigham recommended that the Board analyze "subject by subject, its own examinations in order to discover methods of increasing their validity and reliability." He indicated that the Commission on Scholastic Aptitude Tests had established a model of research that would serve well the entire program of the Board. "If the Board sets up such an experimental machine, it would be the first of its type and any results demonstrably significant should tend to advance the Board's general reputation."¹¹

The Gatekeeper as Usher

As the Board gained confidence in its new examination, it sought ways to broaden that test's impact. The Executive Committee allowed the Commission on the Scholastic Aptitude Test "under suitable conditions to report to a secondary school the result of the test taken by a pupil a year before graduation." The Board saw the test as potentially useful for advising students.

As the test is now divided into two parts, verbal and mathematical, it might be possible for the school to advise its pupils, if they should take the test a year before graduation, whether they should continue to prepare for college, and, if so, whether they should look forward to literary or scientific courses.¹³

11 Ibid., p. 13.

_

¹⁰ Ibid.

¹² Thirty-First Annual Report of the Secretary of the College Board (New York: College Board, 1931), p. 23.

¹³ Ibid.

Such a program was not without its detractors. Brigham himself had strong opinions about such uses of the test; he strongly supported the concept of testing for advising purposes, recommending that

The Board must be prepared to describe the position of individuals in *areas*, as areas, and not with respect to a *line*. The Board examinations might conceivably be extended two or three years below the admissions line and one or two years above this line.¹⁴

He was concerned, however, that the Board would embark on this new venture by simply using the existing SAT without adequate experimentation to develop appropriate instruments:

An organization set up for the sole purpose of collecting tickets at the gate is now asked to show people to their seats. The notion of a general admissions ticket is yielding to the notion of a more exact description of the individual which will make possible his proper placement in definable universes of knowledge.¹⁵

Brigham did not want to use a test designed for one purpose to meet a demand for another use. His advocacy of limited uses for testing may have been a brake on the ambitions of some board members, but, as the SAT grew in importance to the Board, so, too, did Brigham's stature within the Board. The Board was proud of its leading psychologist.

¹⁴ Carl Campbell Brigham "Report To The Executive Committee of the College Entrance Examination Board," March 22, 1933. [ETS Archives Microfilm] p. 9. [Emphasis in original].

¹⁵ "Report of the Associate Secretary," p. 14.

¹⁶ A.B. Crawford, in his report to the Committee on "Testing for Differential Aptitudes," noted that "of course all of our so-called intelligence tests are essentially measures of scholastic aptitude, but many of them (including the American Council Test and the Thorndike C.A.V.D.) yield a single index of capacity only, affording no clue as to individuals' differential abilities, or relative aptitudes for one as compared with another area of study. As Professor Brigham has pointed out, the process of thus striking an average of achievement in different mental stunts tends to obscure the very characteristics which may be most important for guidance purposes." Subcommittee report on "Testing for Differential Aptitudes," Ben D. Wood Papers MSS11.

Brigham's Study of Error

At the September 1932 annual meeting, the College Entrance Examination Board discussed its most recent publication.

Those who are interested in the field of mental measurement will be glad to know of the publication of a book by Professor Carl C. Brigham. . . describing and interpreting the investigations of the Commission on Scholastic Aptitude Tests during the six year period 1926-1931.¹⁷

Brigham's book containing "a detailed study of about one thousand individual test items or questions," was, according to the Board "completely factual and specific, reporting the actual results obtained from the use of the items." 18

In this publication Brigham developed and expressed his theories of testing and education. Brigham saw testing as providing a window on cognitive processes. His introduction to the book noted that he was "interested in all solutions offered as answers to the single item situation as offering the possibility of a systematic study of error." This was in contrast to most psychologists who were looking at the total test as simply a means to develop a linear scale on which people may be given quantitative values. Brigham was interested in more than ranking; he was, in the words of his former colleague, "interested in what made an item work and how people thought about it.¹⁹ Thus he attempted, in his

^{17 &}lt;u>Thirty-Second Annual Report of the Secretary College</u> (New York: College Board, 1932), p. 4.

¹⁸ Ibid.

¹⁹ Brigham, <u>A Study of Error</u>, p. 9 [Brigham's interest in why people make errors extended into his recreation. Chauncey Belknap, a friend of Brigham's from college until Brigham's death, reported that Brigham invented a game called "silly school" for their children. Brigham would ask questions and reward the answer that was most wrong after finding out from the children what they thought made the answer wrong. See: Gary Saretzky "Oral History with Chauncey Belknap" ETS Oral History Program p. 7.

<u>Study of Error</u>, to forge the link between cognitive theory and psychological testing.²⁰

Brigham became noted for his interest in the errors that individuals made when they attempted to solve problems or answer questions.²¹ Ben Wood recalls that the concern with what makes a person miss a particular item "was [Brigham's] unique characteristic."²² As early as 1926, Brigham expressed an interest in the individual test item. "The accuracy of the whole test as a measure of intelligence depends on the merits of each one of the three hundred or more separate hurdles or items contained in the test. Each item stands for a short problem which must be solved correctly."²³

This work anticipated the work of Piaget and the later work of Robert J. Sternberg.²⁴ Brigham was not content to simply find post-hoc correlations between testing instruments and other criteria; he wanted to know what was, in fact, being assessed by a test and even by an individual item.²⁵ He criticized

.

²⁰ For a present day discussion the value of error analysis in educational measurement, see: Robert Glaser, "The Future of Testing," <u>American Psychologist</u>.

²¹ Brigham's interest in the problem solving process by analysis of errors has striking parallels to the work of Jean Piaget. For a brief, clear discussion of Piaget and his relationship to intelligence testing, see: Howard Gardner, <u>Frames of Mind: The Theory of Multiple Intelligences</u> (New York: Basic Books, 1985), pp. 17-21.

²² Ben D. Wood, "Oral History" with David R. Hubin, May 23, 1985, p. 5.

²³ Carl C. Brigham, "Intelligence Tests: The Third of the Present Series of Princeton Lectures by Members of the University Faculty" <u>Princeton Alumni Weekly</u> 26 (May 5, 1926):788.

²⁴ Sternberg is interested in the thought processes that go into solving I.Q. type items. See "A Different Sort of I.Q. Test," in Daniel Goleman, "Rethinking the Value of Intelligence Tests" The New York Times Education Life, (November 9, 1986):24, and Robert J. Sternberg, "The Nature of Mental Abilities" American Psychologist 34 (1979):214-230..

²⁵ Brigham sought information on how all aspects of the test taking experience effected performance. In 1930, while analyzing why a particular type of math item caused a specific type of error, he wanted to determine the effect of speed on math questions. To do this he proposed

those who simply used tests to obtain a ranking. In an insightful illustration of intelligence testing practice preceding and even determining theory, Brigham critically noted in 1932 that

General intelligence seems merely something hypostatized to explain test scores. The conventional practice of a tester of adding all of his scores into a single total score made it necessary to hypostatize a general intelligence and not specific intelligences."²⁶

In other words, if the format of a test and scoring procedures yielded only one number as a score, then only a single concept of intelligence was necessary to explain the results of intelligence tests. The emphasis on the individual item and the cognitive processes that could lead an individual to a particular response moved Brigham into clear opposition, however, to concepts of a single intelligence and genetic determinism.²⁷ This emphasis also affected his views on development of intelligence. He noted, "We gain much from the study of individual performances--we gain nothing from the hypostatization of abilities, powers, faculties, and intelligences which 'explain these performances. . . . If biologists had invented tests, we would have had different abilities hypostatized. On a par with 'general intelligence' is 'heredity' 'ether' and 'a general

an "error study machine" and sought help from the Navy Department as he tried to develop an apparatus that would photograph test takers activities along with electrically operated stop watches. Carl Campbell Brigham to Commander D.E. Cummings, Bureau of Navigation, Navy Department. April 10, 1930. Brigham Folder, ETS Archives. [The evidence suggests that Brigham did, in fact, complete such a machine. Henry Chauncey refers to seeing it "on a visit to Princeton," in a letter to Tom Donlon. See Chauncey to Donlon, December, 1979, Brigham Papers, ETS Archives.

²⁶ Brigham, A Study of Error, p. 28.

²⁷ Brigham became, in fact, recognized as a spokesman for the alternative position. In 1946 Albert Crawford and Paul Burnham note that "the standardized, undiscriminating use of general intelligence tests and the assumptions underlying this procedure were trenchantly attacked by the late Carl C. Brigham in A Study of Error. "Albert B. Crawford and Paul S. Burnham, Forecasting College Achievement: A Survey of Aptitude Tests in Higher Education (New Haven: Yale University Press, 1946), p.78.

education."²⁸ Brigham further expressed the environmentalist's position in asserting that "test findings would not be construed as necessarily revealing unalterable psychological characteristics of the individual, but merely as exposing what is happening to the individual in his culture.²⁹

Although Brigham was thoroughly disposed toward experimentation, he did not promote a closer relationship between testing and experimental psychology per se. In fact, he believed that psychometrics "had become ridiculous only when it has tried to maintain its pseudo-scientific contacts with laboratory psychology by borrowing its phantom formulae." In the thirties Brigham's interests moved away from psychology. Brigham felt it necessary for the data being generated through experimentation with tests to be examined

²⁸ Brigham, <u>A Study of Error</u>, p. 28.

²⁹ Ibid., p. 46. [His allegiance to associates who had not experienced similar changes in attitudes continued. In 1932 Brigham dedicated his <u>Study of Error</u> to Charles Winthrop Gould, the financial sponsor of Brigham's earlier Study of American Intelligence and the man who, as trustee of Coopers Union College had first hired Brigham to study the use of intelligence tests. See: Trustees' Minutes, July 2, 1929, Cooper Union for the Advancement of Science and Art. Cited in Gary D. Saretzky, Research Memorandum: The Sponsor of Carl Campbell Brigham's <u>A Study of American Intelligence: Charles Gould</u>" (Princeton, New Jersey: Educational Testing Service, December 1982), p. 6.

Brigham the environmentalist in 1932 referred to the unrepentant and strident hereditarian as a "Loyal and Devoted Friend, a Constant Source of Inspiration to Those Who Knew Him and Now Cherish His Memory, A Sincere Believer in the Possibility of a Science of Education." See: Brigham, A Study of Error, p. 6. [The degree to which Gould was "a true believer in the Science of Education," as Brigham was coming to view it is unclear. Research remains to be done to determine whether the strident eugenicist Gould had mellowed. Whether Gould and Brigham differed openly by the early 1930s or not, they were fast personal friends. Gary D. Saretzky, ETS Archivist has determined that Gould willed Brigham \$30,000. see: Saretzky, "Sponsor," p. 6.

³⁰ Brigham, A Study of Error, p. 28.

³¹ Although he retained his chair in the Princeton Psychology department he is reputed to have said, in 1937, "I have forgotten all the psychology I ever knew and now know nothing whatsoever about anything. Class of 1912 Record: 25th Year, p. 172; Reproduced in Belknap, Chauncey and Gary D. Saretzky, Oral History with Chauncey Belknap, E.T.S. Oral History Program, June 18, 1980 Educational Testing Service, Princeton, New Jersey, p. 19.

"from another viewpoint which is not psychological." Brigham sought to break the tie between psychology and testing and to tie testing to an emerging science within the general area of education. The evidence suggests that Brigham's new discipline would be concerned with learning and cognition and the measurement of these through testing. Good testing, in Brigham's view could improve education. Brigham, by his own estimation was an "experimental psychologist turned specialist in educational measurement." While not an educator, he developed in A Study of Error a concept of instruction that was based upon the detection and correction of errors. "Detailed information concerning errors should provide the materials for explicit instruction devised to eradicate them."

In 1981 Warren G. Findley, a former associate of Brigham's, wrote, "there was a richness in Brigham's treatment of test content that has, till now, been largely lost." Findley recalls Brigham's emphasis on the connection between test taking activity and cognitive processes. He recounts an exchange in which they discussed this issue:

I recall an occasion with Dr. Brigham when I took the position that a particular test was good because it predicted subsequent performance, pretty much the current definition of predictive validity. Dr. Brigham's response, with a trace of scorn, was that the basic question remained unanswered: 'We are not through until we find out why.' He did not fully convey his fever for this ultimate truth to collaborators in this study. . . we became masters at predicting wrong solutions to math problems without acquiring a sense of the fundamental significance of such insight.³⁶

³² Ibid.

³³ Findley, "Carl C. Brigham Revisited," p. 8.

³⁴ Brigham, A Study of Error, p. 46.

³⁵ Findley, "Carl C. Brigham Revisited," p. 7.

³⁶ Ibid., p. 8.

The Validating Examinations

The Board's examination program in the early thirties continued to be in competition with admissions by certificate programs. Brigham and the Board acknowledged that certificates "from small areas where the schools were known were valuable in deciding a candidate's preparation for college;" however, the Board criticized certificates where they were accepted on a national basis because it was difficult to appraise their value.³⁷

In 1930, the Board proposed that Brigham take on a new responsibility--to develop "validating" examinations that would be similar to the SAT and could be used by institutions that had already admitted students by certificate.³⁸ The Board intended these examinations to be used by colleges that were not already employing the existing examinations of the Board, but were seeking some way to validate the quality of the certificates of their students.³⁹

By spring of 1931, the Board established a separate "Commission on Validating Examinations," chaired by professor Albert Leroy Jones of Columbia. Although Brigham had initial reservations about the new examinations, the lure, for the researcher, of repeat testings that would allow him to test the "reliability" of the SAT and particularly the new mathematics section, overrode his misgivings about the new application.⁴⁰ Brigham quickly took direct charge of the project

³⁷ Minutes of the College Entrance Examination Board April 1, 1931, p. 9, College Board Archives.

³⁸ Minutes of the College Entrance Examination Board October 29, 1930. p. 8. College Board Archives. See also: <u>Thirty-First Annual Report of the Secretary of the College Board</u>," pp. 18-19.

³⁹ Minutes of the College Entrance Examination Board October 28, 1931, p. 6. College Board Archives.

⁴⁰ see: <u>Thirty-First Annual Report of the Secretary of the College Board</u>, p. 24. [The 1931 tests served as second administrations to students who had been "preliminary candidates in

and the new validating examinations became largely a direct reflection of the Scholastic Aptitude Test.⁴¹

The proposed validating examinations included a verbal test, a mathematical test, and a modern language test. The College Board made plans to report the scores of the candidates to the colleges by letter grades---"A indicating a high position, and E a low position."⁴² The scores on the three examinations would be reported separately giving the colleges "evidence of three kinds--the complete secondary school record, the report of the principal, and the validating scores."⁴³ The tests would thus "validate" the information provided by other sources and "there would not be much difficulty in reaching a decision as to a candidate when the records agreed."⁴⁴ The Board did not intend validating examinations to be traditional content examinations, rather they would test "the candidate's ability to do work in the field in question."⁴⁵ Thus the validating examinations were intended to be, in fact, aptitude tests.

In addition to providing colleges with information to validate the information provided by the certificates, the plan for validating examinations was designed to

^{1930.} Brigham and the SATcommission was not initially pleased with the reliability their instrument: "The correlation of the 1930 test with the 1931 test was in the neighborhood of .70 which is very low coefficient of reliability. The Commission is continuing to study the mathematical test but as a precaution advises that the scores be interpreted only for final candidates."]

⁴¹ Minutes of the College Entrance Examination Board October 26, 1931, p. 2. College Board Archives. See also: <u>Thirty-First Annual Report of the Secretary of the College Board</u>," p. 19.

⁴² Ibid., p. 3.

⁴³ Ibid.

⁴⁴ Ibid.

⁴⁵ Minutes of the College Entrance Examination Board October 28, 1931, p. 6. College Board Archives.

give the high schools quantitative information about their students that could provide the authority of testing experts as justifications for decisions of the schools. "It was suggested that secondary schools might find these tests useful in explaining to parents why pupils failed to receive certificates."46

As Brigham worked on the new program, he indicated his faith in his SAT, but he simultaneously saw areas of aptitude that the earlier test did not measure. While he contended that "the Scholastic Aptitude Test measures all grades of attainment," he acknowledged that the SAT was "deficient in measuring the ability to express one's thought in writing."⁴⁷ Therefore his plan for the validating examinations called for the verbal sections of the SAT to be "extended to validate a candidate's ability in English and history."⁴⁸

Upon Brigham's completion of the new examinations, the Board opted to change the name from "validating examinations" to "qualification tests." It then sent letters "to the heads of schools and colleges over a very wide geographical area" inquiring about possible interest in the tests. The response to the new examinations among the secondary schools was, at best, mixed. Although, according to Board records, the private schools were "very enthusiastic" the public schools were critical of the idea. The commission noted that the less than

⁴⁶ Ibid., p.5

⁴⁷ Minutes of the College Entrance Examination Board October 26, 1931, p. 2. College Board Archives.

⁴⁸ Minutes of the College Entrance Examination Board October 28, 1931, p. 6. College Board Archives.

positive reaction could be attributed to the fee.⁴⁹ Seeing what would be at best a limited market, the Board dropped program.⁵⁰

The Pennsylvania Study: Expanding Standardized Testing

Although the Board did not develop further those tests to validate certificates, other agencies and individuals pursued activities that were designed to encourage the wide use of tests. The most notable example of this activity was the so called Pennsylvania Study of Ben D. Wood and William Learned. In contrast to Brigham, who feared premature general application of the new tests, Wood believed that the best way to develop better testing was to use the instruments widely in order to get large samples for norms and correlations. In his efforts to broaden the use of testing, Wood found a natural ally in William S. Learned of the Carnegie Foundation for the Advancement of Teaching. That foundation embraced testing because, in the words of Ellen Lagemann, "in a variety of ways, exact rating of student potential would increase both private and public returns to education."51

Wood's interest in testing and the CFAT's interest in increasing social efficiency coincided as they sought to answer the question of what knowledge a student possesses and acquires as he or she moves through secondary education and into college was the impulse behind a statewide study in Pennsylvania that included an emphasis on testing. The purpose of the Pennsylvania Study soon embraced the question of what the students knew and

⁴⁹ Minutes of the College Entrance Examination Board April 6, 1932, p. 5. College Board Archives.

⁵⁰ Ibid., p. 6. [The program was never resurrected.]

⁵¹ Lagemann, <u>Private Power</u>, p 106.

how much their subject knowledge increased through schooling.⁵² The Foundation was interested in ascertaining, in William Learned and Ben Wood's words, "why are these young people in high school and college?'; 'what is the responsibility of the high school for orienting its pupils with respect to college aims and purposes?" ⁵³ The study designed to answer these questions took ten years and hundreds of thousands of dollars.⁵⁴ In 1928 a large population of Pennsylvania high school seniors and college seniors took tests prepared by Ben D. Wood and the Cooperative Test Service. The resulting report, <u>The Student and His Knowledge</u>, revealed considerable variation among students within the same institutions and among students within the state.⁵⁵

The Pennsylvania Study formed the basis for interest in standardized

Achievement Testing; it clearly suggested that mass administered objective tests

were a more reliable measure of a student's learning and accumulated

knowledge than were high school transcripts. This evidence that test scores

were more reliable than transcripts listing "units", provided support for educators

who favored admission procedures based on testing rather than school

⁵² Ibid., p. 103. See also: Howard J. Savage, <u>Fruit of an Impulse: Forty-Five Years of the Carnegie Foundation 1905-1950</u> (New York: Harcourt Brace and Company, 1953), p 286. For Ben D. Wood's recollections of the study, see Gary D. Saretzky, "Oral History" with Ben D. Wood, February 17, 1978, pp 30-34.

⁵³ William S. Learned and Ben D. Wood, "The Student and His Knowledge," CFAT Bulletin Number 29 (1938), P. 9.

⁵⁴ Lagemann, Private Power, p. 104.

⁵⁵ Ibid., p. 105. The results also revealed wide variation among students who had been exposed to the same amount of study material—a benchmark that came to be known as the "Carnegie Unit." See: 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. p. I-16.

certification.⁵⁶ If the Pennsylvania Study report called into question the usefulness of the high school certificate, it also suggested criticisms of the old form of College Board examination. Unlike the Board's essay examinations, the tests Wood developed for the study were multiple choice and without the element of subjectivity in the actual grading.⁵⁷

The Pennsylvania Study directly promoted the cause of testing in several significant ways. First, the Pennsylvania study led to Ben Wood's creation of the Cooperative Test Service to produce new tests.⁵⁸ Second, the association of Wood and Learned led directly to the proposal of a consolidated non-profit testing agency. Further, the Pennsylvania study provided the impetus for the National Teacher Examinations. Wood recalls that the study showed that many teachers, "were actually below the level of many high school pupil groups in sheer literacy."⁵⁹

Less directly the study promoted testing by initiating William Learned's interest in testing. Learned who would later produce the Graduate Record Examination and lead in the formation of Educational Testing Service began The Pennsylvania Study after attending a discussion of education in Pennsylvania led by deputy state superintendent of instruction James N. Rule. Rule saw an

56 Lagemann, <u>Private Power</u>, p. 107. see also Wechsler, <u>The Qualified Student</u>, p. 247.

⁵⁷ Testing critics would argue that the apparent "objectivity" of the multiple choice examination is misleading. There is subjectivity in the prior determination of what is an acceptable response and in the selection of "attractive distractors." This, however, is a different form of subjectivity than that possible in the open ended essay response type item.

⁵⁸ Lagemann, <u>Private Power</u>, p 109 and John W. Valentine, <u>The College Board and the School Curriculum</u> (New York: College Entrance Examination Board, 1987), p. 46.

⁵⁹ Ben D. Wood and Gary Saretzky, "Oral History," p. 23. [This Oral History provides a useful treatment of Wood's involvement in the NTE as well as other testing ventures.]

emerging crisis in public education and likened "the rising tide of American youth demanding admission into our secondary schools and higher institutions of learning" to "the Yellow Peril." Learned, then a CFAT staff member, sought the speaker out and proposed the comprehensive study of student variability and learning in Pennsylvania. 61

The PEA and Testing

The Pennsylvania Study was not the only activity in the first one-half of the 1930s that both 1) promoted interest in testing and prompted calls for wider use of standardized instruments for both aptitude and achievement assessment and 2) set the stage for the emergence of a national testing agency. By the early 1930s high school educators voiced a new criticism of "repressive" admissions policies. Members of the Progressive Education Association, aware that the country was moving toward an age of universal secondary education, allied with advocates of vocational education to fend off what they viewed as the college's potential to use entrance requirements to distort the function of the high school.⁶² Both Learned and Wood saw the efforts of Progressive Education Association as

⁶⁰ Lagemann, <u>Private Power</u>, p. 101 from <u>Proceedings</u>, Forty First Convention of the Association of Colleges and Secondary Schools of the Middle States and Maryland, November 1927, p. 79.

⁶¹ Valentine, The College Board, p. 46. Valentine contends that Learned was "stimulated by [Rule's] speech. In contrast, Ellen Lagemann reverses the relationship and portrays Rule as influenced by "Learned's views. Lagemann p. 103. For a discussion of Learned's background and his role in the Pennsylvania Study, see: Lagemann, Private Power, pp. 101-108. For Ben Wood's recollections of Learned's involvement, see: Wood and Saretzky, Oral History with Ben Wood, February 17, 1978. pp. 27-35. ETS Archives Oral History Program Educational Testing Service, Princeton, New Jersey. see also: 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. p. I-16.

⁶² Wechsler, <u>The Qualified Student</u>, p. 249. See also, Lagemann, <u>Private Power</u>, p. 112.

a threat. The two men saw testing as "the scientific description of individual differences," primarily for the purposes of sorting. The PEA, in contrast, sought "more pedagogically oriented means of educational evaluation." 63

In 1933 the Progressive Education Association arranged to have students from thirty high schools admitted to colleges without meeting the traditional entrance requirements. Freed from a necessary college preparation curriculum, the thirty schools experimented widely in their course offerings. Initially Learned, who thought that without the possibility of reliance on "units" the schools would rely more on testing, and the CFAT supported the experiment. Learned's goals and those of the PEA seemed compatible; the Carnegie executive sought more precise and objective tests, while the PEA sought curricular reform. The alliance became troubled, however, over the relative priority of the two objectives.

The rift between the PEA and Learned at the CFAT led directly to the impulse to establish a national testing agency.⁶⁷ Learned and Wood became disenchanted with the PEA as they feared that the association would lead in a movement away from "testing and hierarchy as bases for educational organization"; in the words of William Learned, without testing there would be "a

63 Lagemann, Private Power, p. 112.

⁶⁴ Wechsler, <u>The Qualified Student</u>, p. 250. See also, Valentine, <u>The College Board</u>, p.52.

⁶⁵ Lagemann, Private Power, p. 112.

⁶⁶ Ibid., p. 113.

⁶⁷ Ibid.

return to confusion."⁶⁸ Fear of any movement to decrease the significance of testing led Learned directly to propose a national testing agency.

1934 Council on "Problems and Plans"

By the middle of the decade of the 1930s, standardized testing of both achievement and aptitude had grown to the point where the American Council on Education, through its Committee on Problems and Plans in Education, established a special committee for the "Review of the Testing Movement."

George Zook of the Council indicated that the dramatic growth of Ben Wood's Cooperative Test Service--a program supported with A.C.E. funding--prompted the Council's interest in the direction of testing.⁶⁹

The special committee of review included representatives from both secondary and higher education and two of the leaders in educational testing—Carl Brigham and Ben Wood. At five meetings beginning in December of 1935 and ending in April of 1936, the committee heard testimony from over seventy witnesses including educators, psychologists, educational administrators from all levels of education, as well as "people who have used Cooperative Tests." Topics addressed included everything from the broad questions of "the validity of the enterprise," and "the place and function of examinations in the educative process," to such specific issues as the effects of "coaching for examinations,"

⁶⁸ Quoted in Lagemann, <u>Private Power</u>, p. 114, from a memorandum by Howard J. Savage, March 7, 1941.

^{69 &}quot;Report of the Meeting of the Committee on Review of the Testing Situation: Exhibit 3." Ben D. Wood Papers, MSS11, File 126A. E.T.S. Archives

⁷⁰ Ibid.

and "how can tests be used for the selection and training of prospective teachers on all levels?"⁷¹

One general question underlying much of the discussion of the committee involved the possibility of the consolidation of testing agencies. In one of the informal sessions of the "Committee on Review of the Testing Movement," David E. Weglein, Superintendent of the Baltimore city schools proposed: that "In general, there is a distinct need for a central, coordinating agency that will act as a sieve for research." Weglein continued, "the American Council might well facilitate half a dozen research centers, distributed in geographically strategic centers and charged with supplying the central board with research and materials for test construction and test uses." Brigham expressed misgivings about such a direction; the misgivings placed him direct opposition to Wood and presaged later disputes. The SAT's author noted that "the whole enterprise with the Cooperative Test Service as the spearhead was launched without enough study of the job to be done and the complexities, both technical and practical, involved in such a venture." 72 Brigham, who feared that premature use of the tests by a broad spectrum of educators would damage the testing movement, came into direct conflict with Ben D. Wood who believed that wide use of the examinations would provide the data necessary for further development of the testing movement. 73

⁷¹ Ibid., (Clauses I and III and Exhibits I)

^{72 &}quot;Minutes of the Meeting: Committee Discussion" Ben D. Wood Papers MSS11 126B] ETS Archives

⁷³ Wood consistently concerned with building up "norms" Early on, Brigham was not irritated by this. "I wish to thank you very much for your letter of July 16 in which you say that you will be glad to give the College Entrance Examination Board a preemptive control of any of your test one year in advance of their regular publication dates for a nominal legal consideration of one dollar, on the understanding that we would allow you to have the results in order to build up your norms.

Conant's Scholarship Tests

In 1937 two significant events occurred that indicated the acceptance of the decade-old Scholastic Aptitude Test. First, the Board introduced Scholarship tests and second the Carnegie Foundation introduced a Graduate Record Examination. Each of these events, in its own way, would endorse the format and the content of the SAT and that each was a step toward the ultimate formation of the Educational Testing Service a decade later.

In 1937 the Board began a scholarship testing program for Yale,
Princeton, Columbia and Harvard to provide information that could be "used in
conjunction with other information to determine the fitness of a candidate for a
scholarship."⁷⁴ The purpose was not admissions; it was to determine
scholarship awards. This testing program responded to a need expressed by
Harvard's Conant to determine more appropriately those candidates who should
be offered scholarships. The Harvard President sought a way to assure financial
status would not bar anyone from attending Harvard.⁷⁵ Conant sought a
scholarship that could be awarded with the "reasonable certainty that the
recipients would be outstanding."⁷⁶ He saw the possibility of using tests to
assess the academic aptitude of potential scholarship recipients early in the year

__

⁷⁴ John M. Stalnaker, <u>Report on the Scholarship Tests of April 1937</u>, College Entrance Examination Board Research Bulletin No. 8. (New York: College Board, December, 1937), p. 1. and "Minutes of the College Entrance Examination Board," October 27, 1937, p. 7. College Board Archives.

⁷⁵ Henry Chauncey to Mrs. Sharp, September 27, 1961. Brigham Papers, ETS Archives. [This is a memorandum in which Chauncey comments on and makes additions to Matthew Downey's commissioned biography of Brigham.]

 $^{^{76}}$ James B.Conant, $\,\underline{\text{My Several Lives: Memoirs of a Social Inventor}}\,$ (New York: Harper and Row, 1970), p. 418.

so that they could be admitted before they made plans to attend another institution.⁷⁷

The Scholarship Tests of 1937 are significant for three reasons; first, they represented a Carnegie Foundation funded, cooperative effort of the College Entrance Examination Board and other testing agencies. Their introduction thus was a precursor to the formation of the Educational Testing Service.

Second, the Scholarship tests, represented an endorsement of the Scholastic Aptitude Test and its particular scoring formats. And third, they established earlier dates for administration of the SAT, thus placing that examination in a more useful position for admissions decisions.

Scholarship Tests were based on the cooperation of the College Board and the Cooperative Test Service with funding coming directly from the Carnegie Foundation for the Advancement of Teaching. Thus, these tests brought together two significant components of the important changes in testing taking place in the 1930s. The testing battery combined the College Board's Scholastic Aptitude Test with multiple choice achievement tests developed by Ben Wood. This testing battery consisted of first the SAT in the morning followed by objective achievement tests prepared by the Cooperative Test Service of the American Council on Educations.⁷⁹ Each candidate took the SAT along with an achievement test prepared by Ben Wood's Cooperative Test Service.⁸⁰ The

⁷⁷ Valentine, <u>The College Board</u>, p.48.

⁷⁸ Chauncey, "Oral History" p. 11-13 March 31, 1977 E.T.S. Archives

⁷⁹ Ibid., pp. II-8, II-10, II-13, and II-14.

⁸⁰ Wood, "Oral History" February 17, 1978. p. 12.

tests were scored and evaluated separately; the College Board evaluated the SAT and the Educational Records Bureau scored its achievement test.⁸¹

The tests proved popular among colleges with many indicating that such a combination of aptitude and achievement tests would be useful for admission purposes. In April 1937 the Board examined 2,005 candidates competing for scholarships at fourteen colleges. Henry Chauncey worked with the Cooperative Test Service to see whether an admissions test could be developed with content based multiple choice items. In 1940 the College Board introduced its own multiple choice items. The combination of the SAT with the so called achievement tests proved, in the words of Henry Chauncey, "so popular that they made money for the College board right from the start."

In late 1937 John Stalnaker reported that the examinations had been a success. "The coefficients of reliability in general running over .95."83 The success of the Scholarship Tests of 1937 provided further support for those within the Board who wanted to abandon the essay format in favor of "new type" subject area examinations.84

In 1938 the Scholarship Tests took on a new use; they became a part of the admissions process itself. It had become the practice of some Board member institutions to "excuse high-ranking candidates from taking all

⁸¹ Minutes of the College Entrance Examination Board. October 27, 1937, p. 5. College Board Archives.

⁸² Chauncey to Mrs. Sharp, "Brigham Papers," p.2.

⁸³ Minutes of the College Entrance Examination Board. October 27, 1937, p. 6. College Board Archives.

⁸⁴ Chauncey to Mrs. Sharp, "Brigham Papers."

examinations except the Scholastic Aptitude Test."⁸⁵ In response to a number of colleges that wanted tests at an earlier date, in October of 1937, the Board voted to permit these particular candidates to take the Scholastic Aptitude Test in April." Added earlier dates for a specifically defined subpopulation led to a pattern of increasing numbers of administrations of the SAT.

The typical College Board Examinations were given in June. The Board did not want to administer these in April because preparation for the content examinations might "be disturbing to the schools." The change in policy to allow administration of the Scholastic Aptitude Test in April was not, however, without critics. Some correctly saw the move toward allowing the early dates as a step toward an increased reliance on the SAT as opposed to the Board's essay examination. Professor Edward Noyes objected to allowing the SAT administered in April to be used for admissions purposes. He feared that it "would result in a widespread defection of those colleges which still used the June examinations." He stated that "the new tests were not as good for admission purposes as the essay type, and, in the second place, that the framework in secondary education, which was build up by means of the Board's printed examinations, would be affected adversely and secondary education would suffer." Objections such as those of Noyes notwithstanding, the earlier

85 Minutes of the College Entrance Examination Board. October 27, 1937, p. 7.College Board Archives.

⁸⁶ Ibid.

⁸⁷ Minutes of the College Entrance Examination Board. April 12, 1939, p. 9. College Board Archives.

dates for the SAT became increasingly popular and did, in fact, indicate the further ascendency of the SAT.⁸⁸

Founding The GRE--Cloning the SAT

Simultaneously with the development of the Scholarship Tests, another activity funded with Carnegie money brought the Board together with other testing agencies. During the winter of 1935-36, William Learned of the Carnegie Foundation and Ben Wood of the Cooperative Test Service worked with their respective staffs on a two-staged plan to change dramatically testing in America. The plan consisted of first introducing a "Cooperative Graduate Testing Program and then establishing a unified non-profit testing agency." Learned's strategy called for introducing a graduate entrance examination at four prestigious universities that were leaders in graduate education as an initial activity of a new national agency. Carl Brigham's vehement objection to the concept of this new agency killed this part of the plan, but a suggestion by the graduate deans of Columbia, Harvard, Princeton and Yale to unlink the two plans allowed for a graduate admissions examination to proceed.90

In the late 1930s the graduate schools of American universities confronted problems of increasing complexity in selecting their students. Between 1926 and 1938 the number of doctoral degrees more than doubled while the teaching staffs

⁸⁸ The College Board replaced the Scholarship Tests with the Preliminary Scholastic Aptitude Test in the Fall of 1959, an even earlier examination designed for juniors. See: "Preliminary SAT Announced," <u>College Board Review</u> 37 (Winter 1959):1. The PSAT was designed "to fill two needs: the early guidance of college-bound juniors, and the screening of seniors who hope to qualify for scholarship programs which will require the PSAT."

⁸⁹ Savage, Fruit of an Impulse, p. 287.

⁹⁰ Ibid. pp. # For Ben Wood's recollection of the development of the Graduate Record Examination, see: Wood with Saretzky, "Oral History with Ben Wood," p. 45.

in higher education increased less than 50 percent. "With this flood of students the problems of admissions and transfer students had become acute. The officers admitting to the graduate schools at Harvard, Yale, Princeton, and Columbia Universities alone must deal each year with credentials from about five hundred different institutions in America and abroad." The Carnegie Foundation responded to the demand for a new instrument in 1936 with its Graduate Record Examination for use in admissions to graduate liberal arts programs. 92

The Graduate Record Examination reflected a composite of the subject area testing and aptitude testing. Wood and Learned faced difficult questions regarding the content of their achievement tests. The most troublesome issue was whether content area testing should be limited to general knowledge questions that were not specific to a particular discipline. In early 1937 Wood reported to Learned that although the general reaction of the faculty at Columbia was "quite sympathetic to the whole idea" of a graduate admissions tests, faculty members consistently emphasized the need to test at the level of knowledge generally held across disciplines and to "avoid any invasion of the departmental prerogative of exploring the special and higher achievements of pupils in their major fields." Wood, who was generally critical of existing tests and would

^{91 &}quot;Report on Validity Studies and Other Research," (1943). G.R.E. Papers: Box 1-15-2. E.T.S. Archives.

⁹² Gerald V. Lannholm and William B. Schraeder. <u>Predicting Graduate School Success:</u> <u>An Evaluation of the Effectiveness of the Graduate Record Examinations</u> (Princeton, New Jersey: Educational Testing Service, 1951), p.7.

⁹³ Ben D. Wood to William S. Learned, December 21, 1936 E.T.S. Archives, G.R.E. Papers, Box 1. The Decision to avoid testing higher level content knowledge was confirmed in later correspondence between Wood and Learned. See for example: Wood to Learned, January 18, 1937, ETS Archives, GRE Papers, Box. 1. See also: Wood to Learned, December

have welcomed the chance to revise most examinations in higher education, indicated to Learned that, "Of course, the idea of testing the higher levels is a tempting one when we are concerned with entering graduate students; and if by tactful acquiescence now we can gain the confidence and maintain the cooperation of the graduate deans and faculties, we shall win the opportunity in a few years of doing something in the way of improving the special examinations of certain departments--examinations which woefully need improvement, but with which we cannot now safely interfere." The objectives of the achievement portion of the new graduate admissions examinations therefore became:

to measure the general educational background of the entering graduate students. Roughly each unit of the achievement tests should seek to measure the breadth of information of the student and to some extent also measure his ability to solve some of the more general problems in each field of achievement.⁹⁵

As had been the case with the earlier application of intelligence tests among undergraduates in the early 1920s, the original graduate examination was not to have a determining affect on admissions. The original intent was to provide more information on the candidate's

intellectual resources as a prospective scholar: his available working knowledge in various academic fields, and to a certain extent, his powers of comparison, judgment, inference, and problem solving. ⁹⁶

^{21, 1936.} Ben D. Wood Papers. MSS11, File 322. E.T.S. Archives. [In this letter Wood recounts the opposition of foreign language and literature departments to the prospect of their students being tested in the "higher phases of mathematics and natural science."]

⁹⁴ Ibid.

⁹⁵ Wood to Learned, January 18, 1937, ETS Archives, GRE Papers, Box. 1. [Wood proposed that all students be required to take all sections of the examination.]

⁹⁶ 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].

To measure these "powers," Learned sought an aptitude test. The choices regarding the "aptitude" portion of the new examination were less difficult than had been those associated with the content sections. Although Wood initially considered developing new aptitude tests based on the Louis L. Thurstone's groundbreaking work on "primary mental abilities," he and Learned decided to make use of the existing Scholastic Aptitude Test as their measure of aptitude.⁹⁷ In 1936 Learned corresponded with a group including Crawford, and Brigham regarding the format, use and the marketing of a Graduate Record Examination. Brigham declined to become directly involved himself, stressing to Learned that "I hope you also realize how hard it is for me to say 'no' to you personally because of our long association and close friendship, but there comes a time when one's own family has to be taken into consideration."98 Brigham supported, however, Learned's request to use both his SAT and his Mathematics Attainment Test on the new G.R.E.99. The executive committee of the College Board considered this cooperative arrangement important as part of an effort to "make its services available not only to its own constituents but also to other

_

⁹⁷ Wood to Learned, June 15, 1936. Ben D. Wood Papers. MSS11, File 322. E.T.S. ARchives..

⁹⁸ Brigham to Learned, May 15, 1936, G.R.E. Papers: Box 1-1-2. E.T.S. Archives.

⁹⁹ In the words of George Mullins the Board was "very happy to place these tests at your disposal." The Board was unwilling, however, to grant Learned permission to reprint the SATor the Mathematics Attainment Test but proposed that the Board itself print the tests and collect unused copies. This was consistent with the Board's emphasis on security of the SAT For Learned's original request and discussions of the arrangements see: See Learned to Mullins, May 10, 1937 and George W. Mullins to William S. Learned, May 15, 1937. G.R.E. Papers, Learned Correspondence, Box 1-1-3 E.T.S. Archives.

organizations to which it might be useful."¹⁰⁰ Moreover, allowing the Scholastic Aptitude Test to be used as a portion of the new graduate battery allowed "the scores obtained by graduate students . . . [to be] compared with the norms of entering freshmen."¹⁰¹

Without the new testing agency he had sought, Learned faced the problem of finding a home for this most recent innovation; however, the Foundation introduced the Graduate Records examination at the four schools in October 1937. The plan was to "carry out the experiment the first year with only a few institutions with the idea of later expansion if the preliminary experiment seems to indicate good possibilities." Learned, Wood and the four participating schools agreed that the test would not be publicized; however, on November 2, 1937, to the dismay of at least one dean, the New York Times ran an article on the new tests. 103

The process of introducing the new Graduate Record Examination to educators paralleled in significant ways the earlier process of introducing

_

¹⁰⁰ Learned to Chauncey, May 20, 1936. In Ben D. Wood Papers MSS11, Folder 322, E.T.S. Archives. For a record of the Board's reaction Minutes of the College Entrance Examination Board, October 27, 1937. P. 8. College Board archives.

¹⁰¹ Minutes: "Committee on Graduate Examinations" G.R.E. Papers: Box 1-15-9. E.T.S. Archives. Using the SAT for graduate admissions established the basis for later controversies. Some contend that asking candidates to retake essentially the same examination is inappropriate in light of the expense to the candidate. The SAT and the GRE correlate so significantly today that the research by William Ben Schraeder citing the correlation coefficient is now a classified document within the ETS archives. Conversations with that retired ETS researcher during the period between January and May, 1985 indicate that the correlation between the two examinations approximates correlations between repeat takings of the SAT. According to Schraeder, if the high correlation were stressed, markets for the later test would diminish.

¹⁰² Dean George B. Pegram of Columbia to Dean R.G.D. Richardson of Brown. April 28, 1937. G.R.E. Papers, Box 1-1-3. E.T.S. Archives.

¹⁰³ New York Times, November 2, 1937.

"intelligence tests" in higher education during the preceding decade, when colleges and universities first applied the new tests to advising but quickly introduced them into the admissions process. Although Learned stressed the value of the Graduate Record Examination for advising and for the student's self-assessment, he saw that when the G.R.E. was perfected it would be used in a manner that "officers of the graduate schools might have the benefit of the information it affords in selecting their students." The Graduate Record Examination was initially voluntary, and attracting candidates to take the exam was a major concern. In early 1940 Learned wrote to Albert Crawford, "We shall do our best now to make the results as serviceable as possible to each student. I count more than anything on the appeal that the examination will make to the average intelligent mind in the light of his own needs." 105

Also paralleling earlier issues, one of the concerns confronting Learned and Wood was proving the validity of the G.R.E. They recognized that the limited range of grading among graduate institutions made graduate grades quite problematic for validity studies. Scholars also recognized that the students enrolled in graduate programs reflected a very narrow band of ability. 107

¹⁰⁴ William S. Learned to L.P. Eisenhart January 5, 1938. G.R.E. Papers. Box 1-1-5. E.T.S. Archives.

¹⁰⁵ William S. Learned to A.B. Crawford February 19, 1940. G.R.E. Papers Box 1-2-1 Learned Correspondence. E.T.S. Archives.

¹⁰⁶ Stuart Conrad Peterson, "The Measurement and Prediction of Scholastic Achievement on the Graduate Level," Unpublished Ph.D. Dissertation, Department of Psychology, the State University of Iowa. 1943. p. 1.

¹⁰⁷ Stuart Peterson noted the relatively minor differences in performance among graduate students coming from "higher grade" undergraduate programs as opposed to "lower grade" programs. He asserted that the reason for similar performance was that it was probable that only the very talented students from the weaker schools were encouraged to continue whereas, in the better universities institutions that have graduate "propinquity might tend to

Learned favored the correlations with the "informed judgments of two or three men . . . of sufficient acquaintance" with the candidates as the criterion of comparison."108

A total of 1457 graduate students took the test in October and November of 1937 at the four institutions. 109 Brown University joined the four other schools in March 1938. 110 Henry M. Wriston, President of Brown University, a close friend of Learned's and a man who played a significant role in Carnegie Foundation activities, led Brown University into the program. 111 After overcoming faculty opposition that led Learned himself to go to Providence to meet with Brown faculty and convince them of the value of the experiment, Brown joined the original four. 112

To Wriston, however, "it seemed unfortunate that such tests should be applied only to those who go to graduate school;"113 He therefore requested

encourage many students of lesser quality to enter the graduate college." Further, he noted that the motivation required to move to a different school from a weaker college might limit the pool from those weaker colleges. 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. 24.

¹⁰⁸ Learned to Chauncey, January 3, 1940 G.R.E. Papers: Box 1-2-1 Learned Correspondence. E.T.S.

^{109 &}quot;Summary Report of the Graduate Record Examination" ETS Archives GRE File Box 14 Folder # 1.

¹¹⁰ Ibid., See also: Savage, Fruit of an Impulse, p. 290

¹¹¹ Savage, Fruit of an Impulse, p. 242 reports that Wriston was secretary to the Board of the Carnegie Foundation. Dean R.G.D. Richardson of Brown made the initial inquiry regarding Brown's participation in a letter to Dean George B. Pegram of Columbia. April 21, 1937. G.R.E. Papers, Box 1-1-3 E.T.S. Archives.

¹¹² Henry M. Wriston to William Learned February 10, 1938. G.R.E. Papers Box 1-1-5. E.T.S. Archives.

¹¹³ Ibid.

permission from the Carnegie Foundation and approval from his faculty to give it to all seniors. In sweeping, philosophical terms, he described the need for an exit assessment of students leaving Brown. Wriston wrote to the senior class of Brown University on February 10, 1938: "There is an enormous shift in the center of gravity of higher education. That shift is away from the fixed curriculum toward the growth of the individual students." He continued "it is not known to many of you, but it is nonetheless a fundamental fact that Brown is one of the pioneers in this effort. Probably no other university or college in America has so long a continuous record of the study of the intellectual capacities of the entering student." Warning the students that they might be graduating in the midst of a depression, he noted that

This seems, therefore, both from an educational standpoint and from the standpoint of equipping you to get jobs, a desirable time to provide you, if possible, with a comprehensive inventory of the resources which you are carrying away from Brown University. You have now, of course, your academic record, but that academic record has (from the point of view I am discussing) two outstanding defects. In the first place it is a record of what you have known rather than of what you know today. The amount of material you have retained, the amount you now carry as part of your daily equipment, the amount you have available in current resources, is not and cannot be reflected by that record, from the very nature of its organization and of its history."

Wriston explained that "it is only within the last year that advances in the art of taking an intellectual inventory have made possible the realization of any

¹¹⁴ Henry M. Wriston to "The Members of the Senior Class," February 15, 1938. GRE Papers, Box 1, E.T.S. Archives.

¹¹⁵ Ibid., p. 2.

¹¹⁶ Ibid.

such suggestion that I have just made."¹¹⁷ Reassuring the students that the time that the examination took would be worthwhile and that they would not be damaged by invidious comparisons, he wrote, using a new terminology: "This is the application of the theory of relativity to achievement. Let me emphasize again that you are to be scored according to no fixed and arbitrary standard."¹¹⁸

After Brown came other colleges that opted to use the Graduate Record Examination. As had been the case fifteen years earlier as admissions offices started to use the new intelligence tests, the reasons that colleges adopted the Graduate Record Examination varied. In several cases Colleges adopted the examination because of the possibility for enhanced prestige. For example, Union College President Ryan Dixon Fox found that

One great advantage came from the opportunity to compare ourselves with the average of the nine selected colleges. The value in this respect has come in very different ways. A student was in my office only yesterday, apparently laboring under some sort of springtime worry; he was sure that Union College had no intellectual tone at all and was considering transferring to almost any other place. He was so dispirited that I tried to reestablish his confidence in the College and explained to him that our studies showed that we were about the average of a selection off the top layer of American colleges. I showed him the graph with the average line indicated, and it apparently impressed him very deeply."119

By 1940 nine schools representing both private and public institutions --Harvard, Yale, Princeton, Columbia, Brown, Iowa, Minnesota, Wisconsin, and

¹¹⁷ Ibid. [Wriston was referring to the preceding year's work of the Carnegie Foundation and the four schools that had piloted the Graduate examination.

¹¹⁸ Ibid., p. 5.

¹¹⁹ Dixon Ryan Fox to William S. Learned May 6, 1940, G.R.E. Papers, ETS Archives. Box 2.

Michigan-- were using the Graduate Record Examination; other schools were allowed to administer the tests experimentally to their students. Learned and Henry Chauncey launched plans to administer the examination as a part of the graduate admissions process in a "sufficient number of centers to cover the entire country." To do this, Chauncey recommended that Learned follow precisely pattern that had worked for the College Board--"write to the principal of a public or private school in the vicinity, . . . and ask for the name of someone who would be able to administer the examinations and also ask for the use of some room." Chauncey encouraged Learned to approach graduate schools with a direct query: "Will you require the Graduate Record Examination of all candidates for admission if the graduate schools at Yale, Princeton, and Columbia do also?" 124

Ellen Lagemann points out that the introduction of testing into the admissions process for graduate education was important to William Learned and the Foundation because, in Learned's words, "While the college may seek to educate all comers, the graduate or professional school is charged with a duty to select those most likely to render broadly effective intellectual service." 125 Just

¹²⁰ William S. Learned to Henry Chauncey, October 30, 1940. G.R.E. Papers, Box 2. E.T.S. Archives.

¹²¹ Dixon Ryan Fox to William S. Learned May 6, 1940, G.R.E. Papers, ETS Archives. Box 2.

¹²² Henry Chauncey to William S. Learned October 25, 1940 G.R.E. Papers Box 2. E.T.S. Archives.

¹²³ Ibid.

¹²⁴ Ibid.

¹²⁵ William S. Learned, "The Graduate Record Examination: A Memorandum on the General Character and Purpose of the Examination," (1941) Quoted in Lagemann, p. 110.

as changes in the high school had given a boost to college entrance testing, changes in the colleges that were commencing in the 1930s would necessitate entrance testing for graduate programs. The general feeling after the first administration of the G.R.E. was that the experiment had been a success. Truman L. Kelley of Stanford wrote to Learned indicating his enthusiasm and even proposing that a "provision be made for giving the test to this entering law school group." The successful introduction of the new Graduate Records Examination also indicated acceptance of the concept of measuring aptitude in general as well as a specific endorsement of the SAT. Roswell P. Angier, Dean of the Graduate School at Yale, wrote to Learned indicating that Yale was considering altering its application forms for graduate admissions to include a space for the candidate to report "his ratings on any general intelligence tests." 127

As had happened a generation earlier with the post World War I results of the intelligence testing of students beginning higher education, colleges in the early 1940s became conscious of the possible invidious distinctions that could be formed based on the results of the G.R.E. Dean E. Gordon Bill of Dartmouth College wrote to Learned in 1940 indicating that

I have an idea that the results of these examinations for our students will not be particularly exhilarating news for the college because our faculty for years have been geared up to the idea that an ability to use facts rather than a heavy retention of facts was what they were striving for in their teaching. I have known innumerable boys who flunked out of Dartmouth; but who did well at other institutions where memory rather than the practical and logical

¹²⁶ Truman L. Kelley to William S. Learned October 22, 1937. G.R.E. Papers. Box 1-1-4. Learned Correspondence. E.T.S. Archives.

¹²⁷ Roswell P. Angier to William S. Learned April 8, 1940. Learned Correspondence. G.R.E. Papers, E.T.S. Archives.

use of facts was rewarded. In any case, I am certain that the interpretation of the results of these examinations will be a fruitful, even if difficult job. 128

The Aborted Attempt to Found ETS

The introduction of the Graduate Record Examination was quite successful; within a very short time educators accepted it as a tool to meet a need in the graduate admissions process. The other part of Learned and Wood's plan to radically alter testing in the United States--the formation of a national testing agency--did not materialize for over a decade.

In the Spring of 1937 the Carnegie Foundation's Richard M. Gummere talked to Conant about the possibility of merging the College Entrance Examination Board with other testing agencies that were working with college problems. Subsequently, speaking to an assembly of the Educational Records Bureau, in 1937, Conant asked "I have only one final suggestion to make tonight. I hope it may possibly be a constructive one. Would it not be worth while to combine in one organization a number of our present testing and examining agencies?" Conant was not alone in seeking such an innovation; in late 1937 William Learned approached Ben Wood with a plan to develop a "General Examinations Board" to coordinate the development, production, administration, and interpretation of tests. 131 Learned hoped to make testing

¹²⁸ E. Gordon Bill to William S. Learned. April 17, 1940. G.R.E. Papers: Box 1-2-2 E.T.S. Archives.

¹²⁹ Conant, My Several Lives, p 425.

¹³⁰ Ibid.

¹³¹ 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 108. See also, John W. Valentine, The College Board and the School Curriculum (New York: College Entrance Examination Board, 1987), p. 47. [Valentine refers to the Board as the General Education Board.]

more efficient for the nation by eliminating competition and duplication. Ben Wood favored the merger and quickly allied himself with the Learned effort. However, according to Conant, Wood "did not speak for many test makers." With Wood's involvement, the cause of objective testing and of the creation of a unified testing agency became linked. 133

Learned's proposal to the College Board for a consolidation drew a mixed response. Some members, alarmed at the drop in student registration for Board examinations between 1931 and 1937, believed that some change was needed and favored such a plan for consolidation and .¹³⁴ Conant notes that the "last months of 1937 and the first of 1938 were a period of private conferences and exchange of letters among Brigham, W.S. Learned, George W. Mullins (the secretary of the College Board) Ben Wood and George Zook." Ultimately, however, the reaction of the Board to the proposal was not positive.

Several issues shaped the negative reaction of the Board to the proposal. One was the critical assessment that many board members made of the work of the Cooperative Test Service. In the view of many board members, the service that Ben Wood had initiated was flawed by a lack of emphasis on research. Henry Chauncey recalled in a 1977 oral history that the SAT was put together differently from other examinations. He specifically compared it to the subject

133 Conant makes this clear in his recollection of the proposal to form a testing agency in 1937. Conant, My Several Lives, p. 426. Ben D. Wood in his oral history, Saretzky and his oral history with Hubin also links them and displays an antipathy for Carl Campbell Brigham because of the latter's opposition to the consolidation of testing agencies. Wood with Hubin, page [Find the page in Wood interview.]

¹³² Conant, My Several Lives, p. 425.

¹³⁴ Ellen Lagemann, Private Power for the Public Good, p 116.

¹³⁵ Conant, My Several Lives, p. 427.

area tests of the Cooperative Test Service (Ben Wood) and noted "I think the College Board effort and the ETS effort have always been a little more thorough." The antipathy between Wood and the Board was, in fact, long-standing. Ben Wood saw the Cooperative Test Service as a counterbalance to the College Board. "It took me ten years, from 1920 to 1930 roughly, to get enough people in favor of objective tests as against the subjective tests of the College Board." The Board prided itself on its close relationship to the member institutions; Wood, however, reports that the perception of the College Board was that it was a rather elite organization consisting of a relatively small number of upper class colleges and universities.

Certainly the most significant factor shaping the Board's response was the reaction of Carl Campbell Brigham.¹³⁸ Brigham saw two problems in the testing movement as it then existed: first, he saw a trend toward "premature standardization,"--a tendency to develop norms based on large numbers of subjects to lend the appearance of meaning to the tests before experimentation indicated what was being tested. Second he believed that such an agency would diminish even further the inadequate research into basic questions that had to be answered if tests "were to serve the purposes of education."¹³⁹

_

¹³⁶ Chauncey, "Oral History," pp. 2-15 March 31, 1977. Chauncey began his involvement in testing as a user of tests, then pursued statistical analysis; he ultimately became the first president of E.T.S.

¹³⁷ Wood with Saretzky, "Oral History with Ben Wood," p. 1.

¹³⁸ Ellen Lagemann, Private Power for the Public Good, p 117.

¹³⁹ Carl Campbell Brigham, "The Place of Research in a Testing Organization, School and Society, (1937):756-759.

On December 11, 1937, Brigham published an article in <u>School and Society</u> that pointed out that "the present testing movement carries the germs of its own destruction and unless the proposed organization is set to develop a cure for these afflictions it will retard rather than advance education." Brigham the scientist and educator spoke out in fear of the power of marketing. In January, Brigham wrote to Conant because the latter had "gone on record as favoring some sort of consolidation." Brigham's principal objection to the formation of a consolidated testing agency was that "although the word research will be mentioned many times in its charter, the very creation of powerful machinery to do more widely those things that are now being done badly will stifle research, discourage new developments, and establish existing methods, and even existing tests, as the correct ones. 141

Brigham noted that the early work of Binet himself "illustrated the dangers of premature standardization." He then reviewed the history of several examinations. Using the Army Alpha examination as his example He described how a flawed or mediocre test can become accepted and thus perpetuate itself. 143

¹⁴⁰ Carl Campbell Brigham to James B. Conant January 3, 1938. E.T.S. Archives. Brigham File

¹⁴¹ Ibid.

¹⁴² Carl Campbell Brigham to James B. Conant January 3, 1938. E.T.S. Archives. [Copies of the letter were widely distributed. Ben Wood wrote commentary in the margins of his copy. Regarding Brigham's comments on the Binet revision "After twenty years of use and 10 years in revision at a cost of many thousands of dollars, the new-revision is distinctly a revision and basically the same as the 1916 scale."

¹⁴³ Brigham's use of the Alpha as his example of a poor test that became quite marketable is significant given his earlier faith in the test as the foundation for his <u>Study of American Intelligence</u>. In this author's opinion this ringing indictment of the Alpha is further indication of the sincerity of Brigham's rejection of his earlier hereditarian conclusions.

This test is an atrocious one, but it was given over a million times during the war, and it is still used in its original form because so many norms are available. I fail to see any significant or fundamental value in these norms, but they have a great sales appeal.¹⁴⁴

Thus, Brigham feared that the sales or marketing components of a large testing agency would overwhelm the research science components. High Brigham describes how the scientist can become a pawn in a game that he can no longer control. Referring to the American Council on Education Psychological Examination developed by Louis and Thelma Gwinn Thurstone, Brigham speculates that "I am sure that Mr. Thurstone feels that the test is obsolete and should be killed, yet in the face of the popular demand which has been created, he apparently can do nothing." Further, he argues that the

American Council must realize that the test is obsolete, but it is one of their projects which is 'successful' since the test is widely used and they pick up a little income from it. They continue to publish it year after year . . . Here you have the combination of one of the ablest research minds in the testing field (Thurstone) and a supposedly responsible education foundation caught in a web of their own spinning from which they are unable or unwilling to extricate themselves." 147

Ironically, because the drive to form the new agency had become linked with the promotion of multiple choice tests generally, Brigham's initial disputes with William S. Learned and Ben D. Wood placed him in the role of defending the

¹⁴⁴ Carl Campbell Brigham to James B. Conant January 3, 1938. E.T.S. Archives. Brigham File. [Ben Wood's marginal comments specifically objected to Brigham's allegation that norms were used for sales appeal. Wood Papers MSS11 Folder 210.]

¹⁴⁵ Carl Campbell Brigham, "The Place of Research in a Testing Organization, <u>School and Society</u>, (1937):756-759. For Ben Wood's interpretation of Brigham's concerns, see Gary Saretzky and Ben D. Wood, Oral History February 17, 1978, p. 15-17.

¹⁴⁶ Ibid.

¹⁴⁷ Carl Campbell Brigham to James B. Conant January 3, 1938. E.T.S. Archives. Brigham File.

College Board Essay examination from the "objective testers" such as Wood. Brigham criticized those who advocated the new objective tests for their premature codification of instruments. Presaging a theme that he would stress for the rest of his life--the need for experimental flexibility in measurement practice--Brigham observed that the objective test "movement has created new dogmatisms as obstinately defended as the old, and has missed the virility of a frankly experimental attack." 148

Tension between Learned and Brigham was not new. Although the latter had initially approached him to be involved in the Pennsylvania Survey, the two men had since that time consistently differed. In early 1938, Brigham began a letter to Learned, "I am very sorry again to become involved in a controversy with you over the significance of certain aspects of the Pennsylvania study. . ."149 Conant concludes from the perspective of his autobiography that Brigham's "quarrel with Learned was fundamental. They disagreed over what the new tests were designed to show; was it aptitude or achievement?" 150 Wood's interpretation of Brigham's opposition to the merger is that it resulted from Brigham's fear that he would lose his role in the Board, a role that Wood characterizes with mixed metaphor. Wood indicates that Brigham thought he

¹⁴⁸ "Report of the Associate Secretary," p. 11. John Valentine contends that Brigham's use of the word "testers" in a pejorative fashion applied specifically to Learned and Wood.

¹⁴⁹ Carl Campbell Brigham to William S. Learned. January 20, 1938. College Board Archives. [Brigham then continues and comments on "the long painful battle" that he has been engaged in with Learned and Wood.]

¹⁵⁰ Conant, My Several Lives, p. 428.

should "be the czar of the movement" and essentially "Allah" for the "Arabs" who comprised the Board. 151

Although Brigham was certainly the most visible and strident opponent of the consolidation of testing agencies, there were others who shared his concerns. Claude M. Fuess, of Phillips Academy, wrote a congratulatory letter to Brigham after reading "The Place of Research in a Testing Organization." He indicated that

it has appeared at precisely the right moment to do a great deal of good. . . . You have given a warning that no one of these tests is as yet perfect or ever will be, and that we must constantly check up through other fields. I hope that this paper will be given a wide circulation in some national media. It will have all the more effect coming from a man who from the beginning has been sympathetic towards these tests." 152

Conclusion

Brigham had certainly seen the Scholastic Aptitude Test as an instrument that would predict performance. He had also, however, seen the test as an instrument that could help teach us how people think and make errors. For a test to play that latter role, it would have to retain an experimental nature.

In the 1930s Brigham became critical of a concept of a unitary intelligence; influenced by Charles Spearman, he came to see many dimensions to intellect. Moreover, by 1938 Carl Brigham had moved completely into the

¹⁵¹ Oral History February 17, 1978, p. 16. [In response to a follow-up question from Saretzky, Wood implies that Brigham may have resisted any merger of the College Board with other organizations because it would lead to more association with, in Wood's words, the "lower class."

¹⁵² Claude M. Fuess to Brigham, October 23, 1937. College Board Archives. Fuess would later write the fiftieth anniversary commissioned history of Board. In that work, he makes no mention of his earlier support for the opponent of a national testing agency.

¹⁵³ Findley, "Carl C. Brigham Revisited," p. 8.

environmentalist camp in the debate over what shaped intelligence. In the pages of the New York Times in late 1938 a heated debate erupted after Beth L.

Wellman of Iowa State University asserted that "given sufficient time and the right combination of circumstances, children would change in I.Q. in very large amounts." After articles appeared for several months, the Times asked Brigham to comment. Brigham commented, "In my opinion, the war of the I.Q.'s should have ceased somewhere around Nov. 11, 1918. The concept of the I.Q. was red hot in 1914, was cooling rapidly in 1916 and was put out of commission entirely during the war." Brigham then continued: "The original and fallacious concept of the I.Q. was that it reported some mysterious attribute of the individual unrelated to his training but now it is generally conceded that all tests are susceptible to training and to varying degrees of environmental opportunity. The tests measure a result and not its origin." 156

Possibly the most prescient phrase in Brigham's long objection to the formation of an "Educational Testing Service" was his contention that "It is possible to devise tests empirically from any given academic generation, but no

_

¹⁵⁴ Beth L. Wellman, "New Tests Attack Theory of Fixed I.Q." New York Times July 17, 1938, Section D, P.4. Wellman cited three cases where children with initial I.Q.'s of 89, 98, and 98 were tested after intensive training and found to have I.Q.'s of 149, 153, and 167 respectively.

¹⁵⁵ Quoted in W.A. MacDonald, "Brigham Adds Fire to 'War of I.Q.'s': Princeton Expert on Testing Decries Constancy Concept in Intelligence Scores." New York Times, December 4, 1938, p. 10D. [Brigham's dates here are interesting given his 1923 Study of American Intelligence.]

¹⁵⁶ Quoted in W.A. MacDonald, "Brigham Adds Fire to 'War of I.Q.'s': Princeton Expert on Testing Decries Constancy Concept in Intelligence Scores." New York Times, (December 4, 1938):10D.

particular generation must be allowed to cast a blight on future generations."¹⁵⁷ Whether a blight or a blessing, Brigham's Scholastic Aptitude Test was cast upon future generations after the researcher died in 1943. Events of the 1940s set the stage for an ossified instrument to dominate admission into higher education through the last half of the twentieth century.

_

 $^{157\,}$ Carl Campbell Brigham to James B. Conant January 3, 1938. E.T.S. Archives. Brigham File p. 4