

University of Nebraska - Lincoln

DigitalCommons@University of Nebraska - Lincoln

---

Social and Technical Issues in Testing:  
Implications for Test Construction and Usage

Buros-Nebraska Series on Measurement and  
Testing

---

1984

## 7. Aptitude and Achievement Tests: The Curious Case of the Indestructible Strawperson

Anne Anastasi  
*Fordham University*

Follow this and additional works at: <https://digitalcommons.unl.edu/burostestingissues>



Part of the [Educational Assessment, Evaluation, and Research Commons](#)

---

Anastasi, Anne, "7. Aptitude and Achievement Tests: The Curious Case of the Indestructible Strawperson" (1984). *Social and Technical Issues in Testing: Implications for Test Construction and Usage*. 9. <https://digitalcommons.unl.edu/burostestingissues/9>

This Article is brought to you for free and open access by the Buros-Nebraska Series on Measurement and Testing at DigitalCommons@University of Nebraska - Lincoln. It has been accepted for inclusion in Social and Technical Issues in Testing: Implications for Test Construction and Usage by an authorized administrator of DigitalCommons@University of Nebraska - Lincoln.

# 7 Aptitude and Achievement Tests: The Curious Case of the Indestructible Strawperson<sup>1</sup>

Anne Anastasi  
*Fordham University*

In a talk I gave at the 1979 ETS Invitational Conference, I remarked that, if I were suddenly endowed with the appropriate occult powers, I should choose to eliminate certain words from the psychometric vocabulary. Among them were the words *aptitude* and *achievement* (Anastasi, 1980). These terms have led to nearly as much confusion, misinterpretation, and misuse of tests as has the more notorious term *intelligence*. Having been asked once more to discuss the same general topic in 1982, it occurred to me that I might consider *why* the myths that surround these terms are so persistent—and persistent they certainly are.

Let us examine specifically the traditional distinction between aptitude and achievement tests. Aptitudes are typically defined more precisely than intelligence, to designate more narrowly limited cognitive domains. Nevertheless, like intelligence, they have traditionally been contrasted with achievement in testing terminology. This contrast dates from the early days of testing, when it was widely assumed that achievement tests measured the effects of learning, whereas intelligence and aptitude tests measured so-called innate capacity, or potentiality, independently of learning. This approach to testing in turn reflected a simplistic conception of the operation of heredity and environment that prevailed in the 1920s and 1930s. The relevant historical background has been thoroughly examined in a recent book by a science historian, Hamilton Cravens,

---

<sup>1</sup>Paper presented in Invited Symposia: State of the Art Series—Achievement Testing, at the meeting of the American Psychological Association, Washington, D.C., August 1982.

which covers the heredity–environment controversy among American scientists between the two World Wars (Cravens, 1978; see also Anastasi, 1979).

### HISTORICAL ANTECEDENTS

Common misconceptions about the relation between aptitude and achievement tests are highlighted by an index introduced in the 1920s and variously named an achievement quotient or an accomplishment quotient. Both terms having the same initials, this index soon came to be known as the AQ. Its origin is generally attributed to Raymond Franzen (1920, 1922). The AQ could be found by dividing the individual's educational quotient (EQ) by his or her intelligence quotient (IQ). The EQ was the ratio of educational age (EA) to chronological age (CA). The AQ could also be computed more directly by dividing educational age by mental age. The educational age was found by referring the score on an achievement battery to the age norms for that battery. Still another procedure was to use age norms for tests in particular academic subjects, like reading or arithmetic, to find "subject ages" for the individual, and then to average these subject ages to obtain the educational age.

Early textbooks on testing regularly included a discussion of the AQ as a means of evaluating a student's educational performance in relation to that student's intellectual potential—a means of comparing achievement with capacity to learn (Freeman, 1926, 1939; Garrett & Schneck, 1933; Greene, 1941; Lincoln & Workman, 1935; Mursell, 1947). It is interesting to trace the statements about the AQ in texts appearing from the 1920s to the 1940s and early 1950s. Even the earliest discussions called attention to the technical and statistical weaknesses of the AQ as a ratio. The major criticisms fell into two categories: The first category was similar to the now familiar criticisms of the traditional ratio IQ; the second was similar to the equally familiar criticism of grade norms—educational age norms were certainly no better than educational grade norms.

These and other technical criticisms, however, were usually mentioned as limitations, which might be avoided under proper conditions or which should be kept in mind in interpreting results. By the 1940s and early 1950s, the criticisms had become more vigorous. The reader was now told that the AQ *as a technique*, "cannot be recommended" (Mursell, 1947, p. 373), that it has "nearly gone out of use" (Greene, 1941, p. 251), "is in growing disrepute" (Cronbach, 1949, p. 282), and "is now practically extinct" (Anastasi, 1954, p. 463).

Psychological criticisms of the use of AQs, as contrasted to statistical criticisms, are found in some textbooks from the outset. Frank N. Freeman's 1926 book, *Mental Tests*, a widely used text of the period, referred to two unwarranted assumptions: first, that intelligence tests provide a measure of innate capacity independent of training; second, that all educational achievement depends on the

same unitary intellectual capacity (Freeman, 1926, pp. 287–288). These concerns were expressed more mildly and less clearly in other early books. Nevertheless, the same authors who criticized the AQ on either statistical or psychological grounds accepted and even recommended a more general, qualitative, informal procedure for using intelligence test scores in interpreting measures of educational achievement. By midcentury, the AQ itself had in fact disappeared, at least from the major textbooks—but its ghost lingered on.

Closely linked to the AQ is the concept of underachievement and overachievement, which was first introduced in attempts to interpret deviant AQs. If children were performing up to capacity, it was expected that their AQs would be close to 100. Those with AQs under 100 were designated underachievers; those with AQs above 100 were the overachievers. Several writers did express some discomfort with the finding of overachievement as thus measured, because it implied that certain persons were performing above their capacity, which seemed a logical impossibility (e.g., Lincoln & Workman, 1935). Nevertheless, they tried to defend the AQ by attributing values over 100 largely to unreliability of both intelligence tests and educational tests and to inaccuracy of educational age norms. They also suggested that unusually strong interest and motivation might account for a few remaining AQs above 100.

Actually, the question of underachievement and overachievement can be more properly formulated as overprediction and underprediction from the first to the second test (Thorndike, 1963). Such intraindividual differences from one test to another simply reflect the well-known fact that no two tests are perfectly correlated. Of course, this statement is also true of other performance indicators, such as course grades. Among the reasons for the prediction errors in individual cases are not only the unreliability of the measuring instruments but also differences in content coverage, the varied effects of attitudinal and motivational factors on the two measures, and the impact of such intervening experiences as remedial instruction.

It should be noted that underprediction or overprediction will occur regardless of the type of test used. It occurs not only when an intelligence test is used to predict subsequent achievement test performance but also if an achievement test is used to predict subsequent intelligence test performance. Furthermore, the same prediction errors are likely to occur in either direction, whether we estimate scores on the later test from scores on the earlier test, or vice versa. From a practical standpoint, the administration of alternate forms or different levels of an achievement test before and after a course of instruction permits a more accurate analysis of individual accomplishment than does the use of two different tests. To take an extreme example, if achievement in reading comprehension is predicted from a nonverbal intelligence test that is heavily loaded with spatial aptitude, the children with higher spatial than verbal aptitude will look like underachievers, whereas those with higher verbal than spatial aptitude will look like overachievers.

## DEBUNKING VENTURES

Psychology has come a long way since World War I. And some psychometricians have made repeated efforts to exorcise the AQ ghost. That intelligence and aptitude tests are not fundamentally different from achievement tests was illustrated as early as 1927 by Truman L. Kelley. In this connection, Kelley coined the expression “jangle fallacy” to designate the opposite of the “jingle fallacy” whereby things called by the same name are assumed to be the same. Kelley (1927, p. 64) defined the jangle fallacy as “the use of two separate words or expressions covering in fact the same basic situation, but sounding different, as though they were in truth different.” Through an analysis of correlational data, Kelley demonstrated that widely used intelligence tests and achievement batteries overlapped by about 90% (Kelley, 1927, pp. 193–209).

Since that time, other investigators have again reported extensive overlap between these two types of tests (e.g., Coleman & Cureton, 1954; Cronbach, 1970, pp. 284–285). In fact, in some instances, the correlation between intelligence tests and achievement batteries is about as high as the reliability coefficients of each. Over the intervening decades, there have been repeated attempts to dispel the myths and clarify the relation between aptitude and achievement tests. Relevant discussions can be found in the successive editions of widely used textbooks (e.g., Anastasi, 1982; Cronbach, 1970; Thorndike & Hagen, 1977). They can likewise be found in the published reports of conferences devoted wholly or largely to this topic (e.g., DuBois, 1969; D. R. Green, 1974; Schrader, 1980).

In major addresses and papers by psychologists, the terms *aptitude* and *achievement* have been used time and again with precision and with sensitivity to their possible misapprehensions. For example, in his presidential address to the APA Division of Evaluation and Measurement, Bert Green observed that “tests of general verbal and numerical skills are usually called aptitude tests, which is unfortunate since the term *aptitude* seems to suggest an inborn, unchangeable trait. Actually the tests assess developed abilities—skills acquired through years of training and practice with verbal and numerical material [B. F. Green, 1978, p. 669].” Further on, he referred to “the long-range achievement tests we call ‘aptitude tests’ [p. 669].”

It is also enlightening to read what the College Board writes about its Scholastic Aptitude Test (SAT) and its series of achievement tests. In various current publications regularly distributed to students, counselors, and other persons concerned with these tests, the College Board consistently describes the SAT as a measure of developed verbal and mathematical reasoning abilities that are related to successful performance in college (e.g., College Entrance Examination Board, 1981a, 1981e). In a fuller statement, the Board (1981d) adds that the SAT “is not a test of some inborn and unchanging capacity. Scores on the SAT are subject to improvement as educational experience, both in and out of school,

causes these verbal and mathematical abilities to develop." In the same sources, the achievement tests are described as measuring the student's "knowledge and ability to apply that knowledge in specific subject areas." The distinction that emerges is primarily one of breadth versus specificity of test content and of antecedent learning experience.

Following the same trend, Snow (1980) described the SAT as "a test of *extended* or *generalized* achievement designed to be indicative of aptitude for college work, that is, for work requiring broader, deeper, higher, and more elaborate organizations and reorganizations of scholastic learning than that represented *directly* in prior public schooling, or in conventional school achievement tests [pp. 43–44]." At the 1981 ETS Invitational Conference, Christopher Jencks presented a paper in which he discussed the SAT and argued quite convincingly that what the SAT measures is not fundamentally different from what the College Board's achievement tests measure. For many in the audience, these arguments came as no surprise. Jencks went on to suggest, however, that because of widespread misconceptions about the meaning of "aptitude," college-bound high school students do not study the subject matter taught in their high school courses as thoroughly and as earnestly as they otherwise might (Jencks & Crouse, 1982).

Even more recently, the GRE Board has taken decisive action to help dispel the false aptitude–achievement distinctions. In a recent GRE Board Newsletter (1982), it was announced that "effective with October 1982 administrations of the Graduate Record Examinations, the Aptitude Test will become the General Test and the Advanced Tests will be called Subject Tests. The GRE Board approved the name changes to avoid any potential misunderstanding about the purpose of the tests [p. 3]." Viewing the question from a broader perspective, the National Academy of Sciences Committee on Ability Testing, in its recently issued formal report, clearly asserts that both aptitude and achievement tests measure developed abilities, and both serve as indicators of the ability to learn (Wigdor & Garner, 1982, pp. 27, 163).

And so it goes on and on. Still the popular misconceptions persist. These viable misconceptions are especially evident in some of the current popular attacks on testing, particularly on tests such as the SAT and other measures of academic aptitudes. The criticisms follow a monotonously uniform pattern. First comes the false attribution. For example, aptitude tests are supposed to assess innate potential. Second comes disproof, which should be quite easy for such an outrageously irrational and naive statement. Third comes the conclusion: Tests are wrong, bad, and should be abandoned.

This brings me to my subtitle, "The Curious Case of the Indestructible Strawperson." First, the critics set up what in folk language is known as a strawman; but in deference to editorial policies to avoid sexist language, I have renamed it a strawperson. After the many decades of persistent efforts by psychometricians to dispel these misconceptions, anyone who accepts them as the

major premise is certainly building a strawperson. Once the strawperson is up, it is easily demolished, and the demolition carries the tests along with the straw. But the question still remains: *Why* do the misconceptions survive in the first place? *Who* keeps them alive?—certainly not the psychometricians and test constructors.

Actually, the misconceptions survive among the general public and among those test users who are not knowledgeable about either testing or psychology. I would not be so bold as to claim that I have the answer to the indestructibility of my strawperson. But I suggest that one explanation may be found in the desire for magic—the desire for easy answers, quick solutions, and shortcuts. It is the desire to which charlatans have catered across the centuries and which accounts for the popularity of astrology, phrenology, palmistry, and all the other fanciful shortcuts for understanding ourselves and our associates. It is these pseudo-sciences that the first applied psychologists had to compete with. Now that psychology has expanded into the public arena, it is the psychologists themselves who are expected to produce the magic. And, of course, they will be damned if they do and damned if they don't.

### THE CONTINUUM OF DEVELOPED ABILITIES

So much for misconceptions. What do we actually know about the relation between aptitude and achievement tests? We may begin by recalling that *any* psychological test is essentially an objective and standardized measure of a sample of behavior. With regard to cognitive behavior, test scores tell us what the individual is able to do at the time. They do not tell us why individuals perform as they do. To answer that question, we need to know something about each person's experiential background. Both aptitude and achievement tests can be best characterized as tests of developed ability. I first heard this term used in the 1950s by Henry Dyer,<sup>2</sup> in a College Board committee meeting. It was probably an idea ahead of its time and did not then have wide impact. It seems we are now beginning to catch up with it. The term *developed abilities* is appearing with increasing frequency in publications on testing. It will be recalled, too, that the College Board now regularly uses this term to describe the SAT.

---

<sup>2</sup>I am differentiating here between the concept of developed abilities and an experimental battery, the Tests of Developed Abilities, produced by ETS for the College Board in the late 1950s (Anastasi, 1961, pp. 442–443; Dyer, 1954; Dyer & Coffman, 1957). Those tests were eventually abandoned because they proved no more predictive of college success than a combination of the SAT and existing achievement tests in specific fields, while being more costly to prepare, administer, and score and less flexible in their use.

What of the differences between instruments traditionally designated aptitude tests and those designated achievement tests? First, tests of developed abilities do not fall into sharply differentiated categories but rather along a continuum. Both aptitude and achievement tests vary widely among themselves; and those near the center of the continuum overlap to such a degree as to be nearly indistinguishable. Nevertheless, if we arrange the instruments that have traditionally been called aptitude tests and achievement tests in this continuum and strip them of unwarranted assumptions about their nature, we can discern some meaningful and useful differences. A number of such differences have been identified with considerable clarity by several psychometricians, including Lee Cronbach (1970, pp. 281–285), Robert Ebel (1974, p. 316), and Lloyd Humphreys (1974, p. 263), among others. Each formulated the distinction somewhat differently and focused on different aspects of the comparison; but their approaches to the question have much in common. I should like to sum up the distinction between instruments at opposite ends of the continuum under two headings: one pertains to antecedent experience, the other to the use of test scores. From the standpoint of any particular test, we might say that one distinction concerns its past and the other its future.

### Antecedent Experience

The tests traditionally designated *aptitude* tests, at one end of the continuum, differ from those designated *achievement* tests, at the other end, in the degree of precision with which relevant antecedent experience is defined. This does *not* necessarily mean generality or specificity of test content, nor does it imply breadth of transfer effect or of applicability of the instrument. Intelligence tests and educational achievement batteries can be equally broad in content coverage and in the situational scope of their predictive validity. A spatial aptitude test and an achievement test in typewriting can be equally specific and limited in content coverage and in applicability. What I am referring to instead is essentially the *experiential pool* upon which the test constructor draws in formulating test items. This experiential pool is defined with considerable clarity and precision in constructing, let us say, an achievement test in solid geometry, or medieval history, or motor vehicle operation. At the other extreme is a test like the Stanford–Binet, in which the definition specifies little beyond growing up in America in the twentieth century. Broadly oriented educational achievement batteries, which endeavor to dissociate themselves from specific course content, add little to this definition. Their domain of antecedent experience could be defined as growing up and going to school in America in the twentieth century.

I am reminded in this connection of the difference between a learning curve and a growth curve plotted with test scores. The growth curve is a learning curve covering a longer period of time and obtained in the absence of precise knowl-

edge about the independent variables that bring about the observed behavioral changes.

To sum up the first difference, tests of developed ability differ in the degree of precision versus vagueness with which the relevant domain of antecedent experience is defined.

### Use of Test Scores

The second difference concerns the way in which test scores are utilized. It is generally recognized that traditional achievement tests are designed and used primarily to assess current status, whereas traditional aptitude tests are designed and used to predict future performance following a specified learning experience. Typical tests of current status, at one end of this continuum, can be illustrated by a licensing examination (as in obtaining a driver's license), a typing test (as in hiring a secretary), a French test (as in selecting an interpreter), a test to assess the effects of self-study or life experiences (as in credit by examination), and a competency test in so-called basic skills (presumably chosen because they are prerequisite to a wide variety of roles in our contemporary culture).

At the other end of the continuum, we find typical "intelligence" and "aptitude" tests designed particularly for predictive purposes. What can the individual learn—how much and how fast can he or she learn—when put through a particular course of study, educational program, industrial apprenticeship, or other systematic learning experience? I'm sure that at this point many of you are thinking that traditional achievement tests can often serve as effective predictors of future learning. That is certainly true. An achievement test in arithmetic is a good predictor of students' subsequent performance in an algebra class.

We must remember that all tests actually assess current status, whether their purpose is terminal assessment or prediction. Hence it is not surprising that some aptitude tests look very much like achievement tests and vice versa. In fact, some writers (Carroll, 1974; Snow, 1980) have argued for aptitude as a concept or construct, defined as all the characteristics of an individual that predispose him or her to success or failure in new learning or in the performance of some future activity. An aptitude *test*, according to this view, is only one indicant of aptitude; other indicants would include achievement tests, data on prior performance, and information regarding relevant personality and physical characteristics. This definition of aptitude obviously focuses on the predictive use of information about the individual, including current test scores of all sorts.

## MORE ABOUT APTITUDES

Let us take a closer look at the concept of aptitude itself. This, after all, is where myths and excess meanings have accumulated. In discussions of aptitude and achievement tests, it is generally the misconceptions about aptitude that have led

to false distinctions and to misuse of scores. Aptitude, as we have seen, has been identified with the predictive use of tests. Prediction, in turn, has traditionally been linked with the process of selection: Some students are admitted (to college, medical school, or whatever) and others are not; some job applicants are hired and others are not. As a result of several emerging societal changes, selection is beginning to give way to classification. Tests are being used increasingly for such purposes as assisting individuals to choose among courses of study, careers, or other alternative action plans; placing applicants in different jobs for maximal utilization of their individual qualifications; and assessing the prerequisite skills and knowledge of individual students in order to fit instructional programs to specific needs.

In all these contexts, the concept of *diagnostic testing* is coming to replace that of testing for prediction. But the role of tests in diagnosis and prediction is not fundamentally different. In all these situations, appropriate tests should be chosen or constructed in the light of a task analysis of the desired behavior domain—whether identified through an academic curriculum, a career, a particular job, or whatever. To be effective, a predictive or diagnostic test should assess the development of those prerequisite skills and knowledge that the individual needs before taking the next step. Although test content may be drawn from a common pool of experiences shared by the examinee population, the selection of relevant items from that pool should be oriented toward the requirements of the subsequent performance pool. Every test has both this backward and forward reference. The forward reference, however, is especially relevant for tests used to assess one's readiness to advance from where one is to where one wants to go—for instance, into a particular job or educational program.

The concept of aptitude as prerequisite skills and knowledge is exemplified in what Ben Bloom (1976, 1980) calls cognitive entry behaviors and affective entry characteristics. The cognitive entry behaviors include such general skills as reading comprehension, basic quantitative skills, writing competence, logical reasoning processes, and possibly still broader skills such as attention skills and study skills. These are the skills tapped in most scholastic aptitude and academic intelligence tests. Bloom maintains, however, that the more specific cognitive entry behaviors identified as prerequisites for a particular set of learning tasks provide more accurate assessment and are more readily alterable by appropriate instruction. Affective entry characteristics also influence the individual's subsequent learning performance. They include relevant emotional, motivational, and self-concept variables. To some extent, they too can be altered by subsequent instruction adapted to individual needs. Effective instruction requires full information regarding the individual's status upon entry into the instructional program (initial aptitude), as well as clear specification of what is to be learned (achievement goals).

School readiness is another condition associated with the concept of aptitude. It refers essentially to the attainment of prerequisite skills, knowledge, attitudes,

motivations, and other behavioral traits that enable the learner to profit maximally from school instruction. These prerequisites are what Hunt and Kirk (1974) have called the "entry skills" that the child needs to cope with the teaching-learning situation encountered in the first grade. At one time, such readiness was conceived largely in terms of maturation. To be sure, the development of certain minimum physical qualifications facilitates some kinds of learning. Unless children can make the necessary auditory discriminations, they cannot learn to speak by the usual procedures; without the ability for fine motor coordination, they are unable to manipulate a pencil in writing. Most school learning, however, is not so closely linked to sensorimotor development. In the mastery of educational tasks, the importance of prior learning is being increasingly recognized. More and more emphasis is now placed on the hierarchical development of knowledges and skills, whereby the acquisition of simple concepts equips the child for the learning of more complex concepts at any age.

Still another way to conceptualize aptitude and achievement in an educational context is presented by Robert Ebel (1969, 1974). In an incisive analysis of the goals of education, Ebel (1969) concluded that the essence of educational achievement is "command of useful verbal knowledge [p. 66]" and that this objective should be reflected in the construction of educational tests. In order to be meaningful to the individual learner and retrievable when relevant, each new acquisition must be integrated into a coherent structure of knowledge. According to this view, "aptitude for learning consists mainly and essentially of relevant knowledge. . . . What the student has achieved in learning becomes, if it is relevant, his aptitude for further learning" (Ebel, 1974, p. 316). This process cannot occur independently of the subject matter to which it is applied. We do not think content-free thoughts nor develop content-free abilities. The availability of a large, well-organized, and easily retrievable content store is also emerging as a major difference between the performance of expert and novice in such activities as playing chess and solving difficult problems in physics (Glaser, 1981).

The increasing recognition of the importance of the knowledge context of developed abilities is reflected in a recent statement prepared under College Board auspices. The statement concerns the basic academic competencies that college-bound high school students should develop (College Entrance Examination Board, 1981b, 1981c). Following an initial year of intensive discussions by representative groups of educators, a plan was formulated covering both broad developed abilities (called academic competencies) and recommended curricular fields. The list of academic competencies, although defined at a higher academic level, sounds very much like the cognitive entry behaviors described by Bloom. They include developed abilities in reading, writing, listening and speaking, mathematics, reasoning, and studying. A major conclusion was "that acquisition of the competencies and achievement in the curriculum are interdependent—that is, subject matter cannot be mastered without the necessary competencies, and

the competencies cannot be developed in a vacuum without reference to subject-matter content [College Entrance Examination Board, 1981c, p. 10].”

Despite the indestructible strawperson, we have indeed been making steady progress in expanding, clarifying, and refining our understanding of what aptitude and achievement tests measure. Our main problem is still how to communicate this growth in understanding to test users, test takers, and the general public.

## REFERENCES

- Anastasi, A. *Psychological testing* (1st ed.). New York: Macmillan, 1954.
- Anastasi, A. *Psychological testing* (2nd ed.). New York: Macmillan, 1961.
- Anastasi, A. A historian's view of the nature-nurture controversy: Review of H. Cravens, *The triumph of evolution—American scientists and the heredity-environment controversy, 1900–1941*. *Contemporary Psychology*, 1979, 24, 622–623.
- Anastasi, A. Abilities and the measurement of achievement. In W. B. Schrader (Ed.), *Measuring achievement: Progress over a decade*. San Francisco: Jossey-Bass, 1980.
- Anastasi, A. *Psychological testing* (5th ed.). New York: Macmillan, 1982.
- Bloom, B. S. *Human characteristics and school learning*. New York: McGraw-Hill, 1976.
- Bloom, B. S. New directions in educational research: Alterable variables. In W. B. Schrader (Ed.), *Measuring achievement: Progress over a decade*. San Francisco: Jossey-Bass, 1980.
- Carroll, J. B. The aptitude-achievement distinction: The case of foreign language aptitude and proficiency. In D. R. Green (Ed.), *The aptitude-achievement distinction*. Monterey, Calif.: CTB/McGraw-Hill, 1974.
- Coleman, W., & Cureton, E. E. Intelligence and achievement: The “jangle fallacy” again. *Educational and Psychological Measurement*, 1954, 14, 347–351.
- College Entrance Examination Board. *ATP guide for high schools and colleges, 1981–82*. New York: Author, 1981. (a)
- College Entrance Examination Board. *Preparation for college in the 1980s*. New York: Author, 1981. (b)
- College Entrance Examination Board. *Project Equality*. New York: Author, 1981. (c)
- College Entrance Examination Board. *Taking the SAT: A guide to the Scholastic Aptitude Test and the Test of Standard Written English*. New York: Author, 1981. (d)
- College Entrance Examination Board. *Your student report, 1981–82*. New York: Author, 1981. (e)
- Cravens, H. *The triumph of evolution: American scientists and the heredity-environment controversy, 1900–1941*. Philadelphia: University of Pennsylvania Press, 1978.
- Cronbach, L. J. *Essentials of psychological testing*. (1st ed.). New York: Harper, 1949.
- Cronbach, L. J. *Essentials of psychological testing* (3rd ed.). New York: Harper & Row, 1970.
- DuBois, P. H. *Toward a theory of achievement measurement: Proceedings of the 1969 Invitational Conference on Testing Problems*. Princeton, N.J.: Educational Testing Service, 1969.
- Dyer, H. S. *A common philosophy for the Tests of Developed Ability*. Unpublished Memorandum, January 5, 1954.
- Dyer, H. S., & Coffman, W. E. The Tests of Developed Ability. *College Board Review*, 1957, No. 31, 5–10.
- Ebel, R. Knowledge vs. ability in achievement testing. In P. H. DuBois (Ed.), *Toward a theory of achievement measurement*. Princeton, N.J.: Educational Testing Service, 1969.
- Ebel, R. L. The relation of aptitude for learning to achievement in learning. In D. R. Green (Ed.), *The aptitude-achievement distinction*. Monterey, Calif.: CTB/McGraw-Hill, 1974.
- Franzen, R. H. The accomplishment quotient, a school mark in terms of individual capacity. *Teachers College Record*, 1920, 21, 432–440.

- Franzen, R. H. The accomplishment ratio: A treatment of the inherited determinants of disparity in school product. *Teachers College Contributions to Education*, 1922, No. 125.
- Freeman, F. N. *Mental tests: Their history, principles and applications*. Boston: Houghton Mifflin, 1926.
- Freeman, F. N. *Mental tests: Their history, principles, and applications* (rev. ed.). Boston: Houghton Mifflin, 1939.
- Garrett, H. E., & Schneck, M. R. *Psychological tests, methods, and results*. New York: Harper, 1933.
- Glaser, R. The future of testing: A research agenda for cognitive psychology and psychometrics. *American Psychologist*, 1981, 36, 923–936.
- GRE Board Newsletter*. No. 36, February 1982. Graduate Record Examination Board, Princeton, N.J. 08541.
- Green, B. F., Jr. In defense of measurement. *American Psychologist*, 1978, 33, 664–670.
- Green, D. R. (Ed.). *The aptitude-achievement distinction: Proceedings of the Second CTB/McGraw-Hill Conference on Issues in Educational Measurement*. Monterey, Calif.: CTB/McGraw-Hill, 1974.
- Greene, E. B. *Measurements of human behavior*. New York: Odyssey, 1941.
- Humphreys, L. G. The misleading distinction between aptitude and achievement tests. In D. R. Green (Ed.), *The aptitude-achievement distinction*. Monterey, Calif.: CTB/McGraw-Hill, 1974.
- Hunt, J. McV., & Kirk, G. E. Criterion-referenced tests of school readiness. A paradigm with illustrations. *Genetic Psychology Monographs*, 1974, 90, 143–182.
- Jencks, C., & Crouse, J. Should we relabel the SAT—Or replace it? In W. B. Schrader (Ed.), *Measurement, guidance, and program improvement: Proceedings of the 1981 ETS Invitational Conference*. San Francisco: Jossey-Bass, 1982.
- Kelley, T. L. *Interpretation of educational measurements*. Yonkers, N.Y.: World Book Co., 1927.
- Lincoln, E. A., & Workman, L. L. *Testing and the use of test results*. New York: Macmillan, 1935.
- Mursell, J. L. *Psychological testing*. New York: Longmans, Green, 1947.
- Schrader, W. B. (Ed.). *Measuring achievement: Progress over a decade (Proceedings of the 1979 ETS Invitational Conference)*. San Francisco: Jossey-Bass, 1980.
- Snow, R. E. Aptitude and achievement. In W. B. Schrader (Ed.), *Measuring achievement: Progress over a decade*. San Francisco: Jossey-Bass, 1980.
- Thorndike, R. L. *The concepts of over- and under-achievement*. New York: Teachers College Press, 1963.
- Thorndike, R. L., & Hagen, E. *Measurement and evaluation in psychology and education* (4th ed.). New York: Wiley, 1977.
- Wigdor, A. K., & Garner, W. E. (Eds.). *Ability testing: Uses, consequences, and controversies. Part I. Report of the Committee*. Washington, D.C.: National Academy Press, 1982.