

# Chapter 13

## Research on Cognitive, Personality, and Social Psychology: I

Lawrence J. Stricker

Several months before ETS's founding in 1947, Henry Chauncey, its first president, described his vision of the research agenda:

Research must be focused on objectives not on methods (they come at a later stage). Objectives would seem to be (1) advancement of test theory & statistical techniques (2) refinement of description & measurement of intellectual & personal qualities (3) development of tests for specific purposes (a) selection (b) guidance (c) measurement of achievement. (Chauncey 1947, p. 39)

By the early 1950s, research at ETS on intellectual and personal qualities was already proceeding. Cognitive factors were being investigated by John French (e.g., French 1951b), personality measurement by French, too (e.g., French 1952), interests by Donald Melville and Norman Frederiksen (e.g., Melville and Frederiksen 1952), social intelligence by Philip Nogee (e.g., Nogee 1950), and leadership by Henry Ricciuti (e.g., Ricciuti 1951). And a major study, by Frederiksen and William Schrader (1951), had been completed that examined the adjustment to college by some 10,000 veterans and nonveterans.

Over the years, ETS research on those qualities has evolved and broadened, addressing many of the core issues in cognitive, personality, and social psychology. The emphasis has continually shifted, and attention to different lines of inquiry has waxed and waned, reflecting changes in the *Zeitgeist* in psychology, the composition of the Research staff and its interests, and the availability of support, both external and from ETS. A prime illustration of these changes is the focus of research at ETS and in the field of psychology on level of aspiration in the 1950s, exemplified by the ETS studies of Douglas Schultz and Henry Ricciuti (e.g., Schultz and

---

This chapter was originally published in 2013 as a research report in the ETS R&D Scientific and Policy Contributions Series.

L.J. Stricker (✉)  
Educational Testing Service, Princeton, NJ, USA  
e-mail: [lstricker@ets.org](mailto:lstricker@ets.org)

Ricciuti 1954), and on emotional intelligence 60 years later, represented by ETS investigations by Richard Roberts and his colleagues (e.g., Roberts et al. 2006).

What has been studied is so varied and so substantial that it defies easy encapsulation. Rather than attempt an encyclopedic account, a handful of topics that were the subjects of extensive and significant ETS research, very often in the forefront of psychology, will be discussed. In this chapter, the topics in cognitive psychology are the structure of abilities; in personality psychology, response styles, and social and emotional intelligence; and in social psychology, prosocial behavior and stereotype threat. Motivation is also covered. The companion chapter (Kogan, Chap. 14, this volume) discusses other topics in cognitive psychology (creativity), personality psychology (cognitive styles, kinesthetic after effects), and social psychology (risk taking).

### 13.1 The Structure of Abilities

Factor analysis has been the method of choice for mapping the ability domain almost from the very beginning of ability testing at the turn of the twentieth century. Early work, such as Spearman's (1904), focused on a single, general factor ("g"). But subsequent developments in factor analytic methods in the 1930s, mainly by Thurstone (1935), made possible the identification of multiple factors. This research was closely followed by Thurstone's (1938) landmark discovery of seven primary mental abilities. By the late 1940s, factor analyses of ability tests had proliferated, each analysis identifying several factors. However, it was unclear what factors were common across these studies and what were the best measures of the factors.

To bring some order to this field, ETS scientist John French (1951b) reviewed all the factor analyses of ability and achievement that had been conducted through the 1940s. He identified 59 different factors from 69 studies and listed tests that measured these factors. (About a quarter of the factors were found in a single study, and the same fraction did not involve abilities.)

This seminal work underscored the existence of a large number of factors, the importance of replicable factors, and the difficulty of assessing this replicability in the absence of common measures in different studies. It eventuated in a major ETS project led by French—with the long-term collaboration of Ruth Ekstrom and with the guidance and assistance of leading factor analysts and assessment experts across the country—that lasted almost two decades. Its objectives were both (a) substantive—to identify well-established ability factors and (b) methodological—to identify tests that define these factors and hence could be included in new studies as markers to aid in interpreting the factors that emerge. The project evolved over three stages.

At the first conference in 1951, organized by French, chaired by Thurstone, and attended by other factor analysts and assessment experts, French (1951a) reported that (a) 28 factors appeared to be reasonably well established, having been found in at least three different analyses; and (b) 29 factors were tentatively established, appearing with "reasonable clarity" (p. 8) in one or two analyses. (Several factors in each set were not defined by ability measures.) Committees were formed to verify

the factors and identify the tests that defined them. Sixteen factors and three corresponding marker tests per factor were ultimately identified (French 1953, 1954). The 1954 Kit of Selected Tests for Reference Aptitude and Achievement Factors contained the tests selected to define the factors, including some commercially published tests (French 1954).

At a subsequent conference in 1958, plans were formulated to evaluate 46 replicable factors (including those already in the 1954 Kit) that were candidates for inclusion in a revised Kit and, as far as possible, develop new tests in place of the published tests to obviate the need for special permission for their use and to make possible a uniform format for all tests in the Kit (French 1958). Again, committees evaluated the factors and identified marker tests. The resulting 1963 Kit of Reference Tests for Cognitive Factors (French et al. 1963) had 24 factors, along with marker tests. Most of the tests were created for the 1963 Kit, but a handful were commercially published tests.

At the last conference, in 1971, plans were made for ETS staff to appraise existing factors and newly observed ones and to develop ETS tests for all factors (Harman 1975). The recent literature was reviewed and studies of 12 new factors were conducted to check on their viability (Ekstrom et al. 1979). The Kit of Factor-Referenced Cognitive

**Table 13.1** Factors and sample marker tests in Kit of Factor-Referenced Cognitive Tests, 1976

Factor	Marker test
General reasoning	Necessary arithmetic operations
Induction	Letter sets
Logical reasoning	Nonsense syllogisms
Integrative processes	Following directions
Verbal comprehension	Vocabulary test 1
Number facility	Addition
Spatial orientation	Card rotations
Visualization	Paper folding
Spatial scanning	Maze tracing
Perceptual speed	Number comparison
Flexibility of closure	Hidden figures
Speed of closure	Gestalt completion
Verbal closure	Scrambled words
Memory span	Auditory number span
Associative memory	First and last names
Visual memory	Map memory
Figural Fluency	Ornamentation
Expressional Fluency	Arranging words
Word Fluency	Word beginnings
Associational Fluency	Opposites
Ideational Fluency	Thing categories
Flexibility of use	Different uses
Figural flexibility	Toothpicks

*Note:* Adapted from *Essentials of Psychological Testing* (5th ed.), by L. J. Cronbach, (1990), New York: Harper & Row

Tests, 1976 (Ekstrom et al. 1976) had 23 factors and 72 corresponding tests. The factors and sample marker tests appear in Table 13.1, as roughly grouped by Cronbach (1990).

Research and theory about ability factors has continued to advance in psychology since the work on the Kit ended in the 1970s, most notably Carroll's (1993) identification of 69 factors from a massive reanalysis of extant, factor-analytic studies through the mid-1980s, culminating in his three-stratum theory of cognitive abilities. Nonetheless, the Kit project has had a lasting impact on the field. The various Kits were, and are, widely used in research at ETS and elsewhere. The studies include not only factor analyses of large sets of tests that use a number from the Kit to define factors (e.g., Burton and Fogarty 2003), in keeping with its original purpose, but also many small-scale experiments and correlational investigations that simply use a few Kit tests to measure specific variables (e.g., Hegarty et al. 2000). It is noteworthy that versions of the Kit have been cited 2308 times through 2016, according to the *Social Science Citation Index*.

## 13.2 Response Styles

Response styles are

... expressive consistencies in the behavior of respondents which are relatively enduring over time, with some degree of generality beyond a particular test performance to responses both in other tests and in non-test behavior, and usually reflected in assessment situations by consistencies in response to item characteristics other than specific content. (Jackson and Messick 1962a, p. 134)

Although a variety of response styles has been identified on tests, personality inventories, and other self-report measures, the best known and most extensively investigated are acquiescence and social desirability. Both have a long history in psychological assessment but were popularized in the 1950s by Cronbach's (1946, 1950) reviews of acquiescence and Edwards's (1957) research on social desirability. As originally defined, acquiescence is the tendency for an individual to respond *Yes, True*, etc. to test items, regardless of their content; social desirability is the tendency to give a socially desirable response to items on self-report measures, in particular.

ETS scientist Samuel Messick and his longtime collaborator at Pennsylvania State University and the University of Western Ontario, Douglas Jackson, in a seminal article in 1958 redirected this line of work by reconceptualizing response sets as response styles to emphasize that they represent consistent individual differences not limited to reactions to a particular test or other measure. Jackson and Messick underscored the impact of response styles on personality and self-report measures generally, throwing into doubt conventional interpretations of the measures based on their purported content:

In the light of accumulating evidence it seems likely that the *major common factors in personality inventories of the true-false or agree-disagree type*, such as the MMPI and the California Personality Inventory, *are interpretable primarily in terms of style rather than specific item content*. (original italics; Jackson and Messick 1958, p. 247)

Messick, usually in collaboration with Jackson, carried out a program of research on response styles from the 1950s to the 1970s. The early work documented acquiescence on the California F scale, a measure of authoritarianism. But the bulk of the research focused on acquiescence and social desirability on the MMPI. In major studies (Jackson and Messick 1961, 1962b), the standard clinical and validity scales (separately scored for the true-keyed and false-keyed items) were factor analyzed in samples of college students, hospitalized mental patients, and prisoners. Two factors, identified as acquiescence and social desirability, and accounting for 72–76% of the common variance, were found in each analysis. The acquiescence factor was defined by an acquiescence measure and marked by positive loadings for the true-keyed scales and negative loadings for the false-keyed scales. The social desirability factor's loadings were closely related to the judged desirability of the scales.

A review by Fred Damarin and Messick (Damarin and Messick 1965; Messick 1967, 1991) of factor analytic studies by Cattell and his coworkers (e.g., Cattell et al. 1954; Cattell and Gruen 1955; Cattell and Scheier 1959) of response style measures and performance tests of personality that do not rely on self-reports, suggested two kinds of acquiescence: (a) uncritical agreement, a tendency to agree; and (b) impulsive acceptance, a tendency to accept many characteristics as descriptive of the self. In a subsequent factor analysis of true-keyed and false-keyed halves of original and reversed MMPI scales (items revised to reverse their meaning), two such acquiescence factors were found (Messick 1967).

The Damarin and Messick review (Damarin and Messick 1965; Messick 1991) also suggested that there are two kinds of socially desirable responding: (a) a partially deliberate bias in self-report and (b) a nondeliberate or autistic bias in self-regard. This two-factor theory of desirable responding was supported in later factor analytic research (Paulhus 1984).

The findings from this body of work led to the famous *response style controversy* (Wiggins 1973). The main critics were Rorer and Goldberg (1965a, b) and Block (1965). Rorer and Goldberg contended that acquiescence had a negligible influence on the MMPI, based largely on analyses of correlations between original and reversed versions of the scales. Block questioned the involvement of both acquiescence and social desirability response styles on the MMPI, based on his factor analyses of MMPI scales that had been balanced in their true-false keying to minimize acquiescence and his analyses of the correlations between a measure of the putative social desirability factor and the Edwards Social Desirability scale. These critics were rebutted by Messick (1967, 1991) and Jackson (1967). In recent years this controversy has reignited, focusing on whether response styles affect the criterion validity of personality measures (e.g., McGrath et al. 2010; Ones et al. 1996).

This work has had lasting legacies for both practice and research. Assessment specialists commonly recommend that self-report measures be balanced in keying (Hofstee et al. 1998; McCrae et al. 2001; Paulhus and Vazire 2007; Saucier and Goldberg 2002), and most recent personality inventories (Jackson Personality Inventory, NEO Personality Inventory, Personality Research Form) follow this practice. It is also widely recognized that social desirability response style is a potential threat to the validity of self-report measures and needs to be evaluated (American

Educational Research Association et al. 1999). Research on this response style continues, evolved from its conceptualization by Damarin and Messick (Damarin and Messick 1965; Messick 1991) and led by Paulhus (e.g., Paulhus 2002).

### 13.3 Prosocial Behavior

Active research on positive forms of social behavior began in psychology in the 1960s, galvanized at least in part by concerns about public apathy and indifference triggered by the famous Kitty Genovese murder (a New York City woman killed reportedly while 38 people watched from their apartments, making no efforts to intervene;<sup>1</sup> Latané and Darley 1970; Manning et al. 2007). This *prosocial behavior*, a term that ETS scientist David Rosenhan (Rosenhan and White 1967) and James Bryan (Bryan and Test 1967), an ETS visiting scholar and faculty member at Northwestern University, introduced into the social psychological literature to describe all manner of positive behavior (Wispé 1972), has many definitions. Perhaps the most useful is Rosenhan's (1972):

...while the bounds of prosocial behavior are not rigidly delineated, they include these behaviors where the emphasis is ...upon "concern for others." They include those acts of helpfulness, charitability, self-sacrifice, and courage where the possibility of reward from the recipient is presumed to be minimal or non-existent and where, on the face of it, the prosocial behavior is engaged in for its own end and for no apparent other. (p. 153)

Rosenhan and Bryan, working independently, were at the forefront of research on this topic in a short-lived but intensive program of research at ETS in the 1960s. The general thrust was the application of social learning theory to situations involving helping and donating, in line with the prevailing *Zeitgeist*. The research methods ran the gamut from surveys to field and laboratory experiments. And the participants included the general public, adults, college students, and children.

Rosenhan (1969, 1970) began by studying civil rights activists and financial supporters. They were extensively interviewed about their involvement in the civil rights movement, personal history, and ideology. The central finding was that fully committed activists had close affective ties with parents who were also fully committed to altruistic causes.

Rosenhan and White (1967) subsequently put this result to the test in the laboratory. Children who observed a model donate to charity and then donated in the model's presence were more likely to donate when they were alone, suggesting that both observation and rehearsal are needed to internalize norms for altruism. However, these effects occurred whether or not the children had positive or negative interactions with the model.

In a follow-up study, White (1972) found that children's observations of the model per se did not affect their subsequent donations; the donations were influ-

---

<sup>1</sup>Subsequent inquiries cast doubt on the number of witnesses and on whether any intervened (Manning et al. 2007).

enced by whether the children contributed in the model's presence. Hence, rehearsal, not observation, was needed to internalize altruistic norms. White also found that these effects persisted over time.

Bryan also carried out a mix of field studies and laboratory experiments. Bryan and Michael Davenport (Bryan and Davenport 1968), using data on contributions to *The New York Times* 100 Neediest Cases, evaluated how the reasons for being dependent on help were related to donations. Cases with psychological disturbances and moral transgressions received fewer donations, presumably because these characteristics reduce interpersonal attractiveness, specifically, likability; and cases with physical illnesses received more contributions.

Bryan and Test (1967) conducted several ingenious field experiments on the effects of modeling on donations and helping. Three experiments involved donations to Salvation Army street solicitors. More contributions were made after a model donated, and whether or not the solicitor acknowledged the donation (potentially reinforcing it). Furthermore, more White people contributed to White than Black solicitors when no modeling was involved, suggesting that interpersonal attraction—the donors' liking for the solicitors—is important. In the helping experiment, more motorists stopped to assist a woman with a disabled car after observing another woman with a disabled car being assisted.

Bryan and his coworkers also carried out several laboratory experiments about the effects of modeling on helping by college students and donations by children. In the helping study, by Test and Bryan (1969), the presence of a helping model (helping with arithmetic problems) increased subsequent helping when the student was alone, but whether the recipient of the helping was disabled and whether the participant had been offered help (setting the stage for reciprocal helping by the participant) did not affect helping.

In Bryan's first study of donations (Midlarsky and Bryan 1967), positive relationships with the donating model and the model's expression of pleasure when the child donated increased children's donations when they were alone. In a second study, by Bryan and Walbek (1970, Study 1), the presence of the donating model affected donations, but the model's exhortations to be generous or to be selfish in making donations did not.

Prosocial behavior has evolved since its beginnings in the 1960s into a major area of theoretical and empirical inquiry in social and developmental psychology, and sociology (e.g., see the review by Penner et al. 2005). The work has broadened over the years to include such issues as its biological and genetic causes, its development over the life span, and its dispositional determinants (demographic variables, motives, and personality traits). The focus has also shifted from the laboratory experiments on mundane tasks to investigations in real life that concern important social issues and problems (Krebs and Miller 1985), echoing Rosenhan's (1969, 1970) civil rights study at the very start of this line of research in psychology some 50 years ago.

## 13.4 Social and Emotional Intelligence

Social intelligence and its offshoot, emotional intelligence, have a long history in psychology, going back at least to Thorndike's famous *Harper's Monthly Magazine* article (Thorndike 1920) that described social intelligence as "the ability to understand and manage men and women, boys and girls—to act wisely in human relations" (p. 228). The focus of this continuing interest has varied over the years from accuracy in judging personality in the 1950s (see the review by Cline 1964); to skill in decoding nonverbal communication (see the review by Rosenthal et al. 1979) and understanding and coping with the behavior of others (Hendricks et al. 1969; O'Sullivan and Guilford 1975) in the 1970s; to understanding and dealing with emotions from the 1990s to the present. This latest phase, beginning with a seminal article by Salovey and Mayer (1990) on emotional intelligence and galvanized by Goleman's (1995) popularized book, *Emotional Intelligence: Why It Can Matter More Than IQ*, has engendered enormous interest in the psychological community and in the public.

ETS research on this general topic started in 1950 but until recently was scattered and modest, limited to scoring and validating situational judgment tests of social intelligence. These efforts included studies by Norman Cliff (1962), Philip Nogee (1950), and Lawrence Stricker and Donald Rock (1990). Substantial work on emotional intelligence at ETS by Roberts and his colleagues began more recently. They have conducted several studies on the construct validity of maximum-performance measures of emotional intelligence. Key findings are that the measures define several factors and relate moderately with cognitive ability tests, minimally with personality measures, and moderately with college grades (MacCann et al. 2010, 2011; MacCann and Roberts 2008; Roberts et al. 2006).

In a series of critiques, reviews, and syntheses of the extant research literature, Roberts and his colleagues have attempted to bring order to this chaotic and burgeoning field marked by a plethora of conceptions, "conceptual and theoretical incoherence" (Schulze et al. 2007, p. 200), and numerous measures of varying quality. These publications emphasize the importance of clear conceptualizations, adherence to conventional standards in constructing and validating measures, and the need to exploit existing measurement approaches (e.g., MacCann et al. 2008; Orchard et al. 2009; Roberts et al. 2005, 2008, 2010; Schulze et al. 2007).

More specifically, the papers make these major points:

1. In contrast to diffuse conceptions of emotional intelligence (e.g., Goleman 1995), it is reasonable to conceive of this phenomenon as consisting of four kinds of cognitive ability, in line with the view that emotional intelligence is a component of intelligence. This is the Mayer and Salovey (1997) four-branch model that posits these abilities: perceiving emotions, using emotions, understanding emotions, and managing emotions.
2. Given the ability conception of emotional intelligence, it follows that appropriate measures assess maximum performance, just like other ability tests. Self-report measures of emotional intelligence that appraise typical performance are inap-



appropriate, though they are very widely used. It is illogical to expect that people lacking in emotional intelligence would be able to accurately report their level of emotional intelligence. And, empirically, these self-report measures have problematic patterns of relations with personality measures and ability tests: substantial with the former but minimal with the latter. In contrast, maximum performance measures have the expected pattern of correlations: minimal with personality measures and substantial with ability tests.

3. Maximum performance measures of emotional intelligence have unusual scoring and formats, unlike ability tests, that limit their validity. Scoring may be based on expert judgments or consensus judgments derived from test takers' responses. But the first may be flawed, and the second may disadvantage test takers with extremely high levels of emotional intelligence (their responses, though appropriate, diverge from those of most test takers). Standards-based scoring employed by ability tests obviates these problems. Unusual response formats include ratings (e.g., presence of emotion, effectiveness of actions) rather than multiple choice, as well as instructions to predict how the test taker would behave in some hypothetical situation rather than to identify what is the most effective behavior in the situation.
4. Only one maximum performance measure is widely used, the Mayer-Salovey-Caruso Emotional Intelligence Test (Mayer et al. 2002). Overreliance on a single measure to define this phenomenon is "a suboptimal state of affairs" (Orchard et al. 2009, p. 327). Other maximum performance methods, free of the measurement problems discussed, can also be used. They include implicit association tests to detect subtle biases (e.g., Greenwald et al. 1998), measures of ability to detect emotions in facial expressions (e.g., Ekman and Friesen 1978), inspection time tests to assess how quickly different emotions can be distinguished (e.g., Austin 2005), situational judgment tests (e.g., Chapin 1942), and affective forecasting of one's emotional state at a future point (e.g., Hsee and Hastie 2006).

It is too early to judge the impact of these recent efforts to redirect the field. Emotional intelligence continues to be a very active area of research in the psychological community (e.g., Mayer et al. 2008).

### 13.5 Stereotype Threat

Stereotype threat is a concern about fulfilling a negative stereotype regarding the ability of one's group when placed in a situation where this ability is being evaluated, such as when taking a cognitive test. These negative stereotypes exist about minorities, women, the working class, and the elderly. This concern has the potential for adversely affecting performance on the ability assessment (see Steele 1997). This phenomenon has clear implications for the validity of ability and achievement tests, whether used operationally or in research.

Stereotype threat research began with the seminal experiments by Steele and Aronson (1995). In one of the experiments (Study 2), for instance, they reported that the performance of Black research participants on a verbal ability test was lower when it was described as diagnostic of intellectual ability (priming stereotype threat) than when it was described as a laboratory task for solving verbal problems; in contrast, White participants' scores were unaffected.

Shortly after the Steele and Aronson (1995) work was reported, Walter McDonald, then director of the *Advanced Placement Program*® (AP®) examinations at ETS, commissioned Stricker to investigate the effects of stereotype threat on the AP examinations, arguing that ETS would be guilty of "educational malpractice" if the tests were being affected and ETS ignored it. This assignment eventuated in a program of research by ETS staff on the effects of stereotype threat and on the related question of possible changes that could be made in tests and test administration procedures.

The initial study with the AP Calculus examination and a follow-up study (Stricker and Ward 2004), with the Computerized Placement Tests (CPTs, now called the *ACCUPLACER*® test), a battery of basic skills tests covering reading, writing, and mathematics, were stimulated by a Steele and Aronson (1995, Study 4) finding. These investigators observed that the performance of Black research participants on a verbal ability test was depressed when asked about their ethnicity (making their ethnicity salient) prior to working on the test, while the performance of White participants was unchanged. The AP examinations and the CPTs, in common with other standardized tests, routinely ask examinees about their ethnicity and gender immediately before they take the tests, mirroring the Steele and Aronson experiment. The AP and CPTs studies, field experiments with actual test takers, altered the standard test administration procedures for some students by asking the demographic questions after the test and contrasted their performance with that of comparable students who were asked these questions at the outset of the standard test administration. The questions had little or no effect on the test performance of Black test takers or the others—Whites, Asians, women, and men—in either experiment. These findings were not without controversy (Danaher and Crandall 2008; Stricker and Ward 2008). The debate centered on whether the AP results implied that a substantial number of young women taking the test were adversely affected by stereotype threat.

Several subsequent investigations also looked at stereotype threat in field studies with actual test takers, all the studies motivated by the results of other laboratory experiments by academic researchers. Alyssa Walters et al. (2004) examined whether a match in gender or ethnicity between test takers and test-center proctors enhanced performance on the *GRE*® General Test. This study stemmed from the Marx and Roman (2002) finding that women performed better on a test of quantitative ability when the experimenter was a woman (a competent role model) while the experimenter's gender did not affect men's performance. Walters et al. reported that neither kind of match between test takers and their proctors was related to the test takers' scores for women, men, Blacks, Hispanics, or Whites.

Michael Walker and Brent Bridgeman (2008) investigated whether the stereotype threat that may affect women when they take the *SAT*® Mathematics section

spills over to the Critical Reading section, though a reading test should not ordinarily be prone to stereotype threat for women (there are no negative stereotypes about their ability to read). The impetus for this study was the report by Beilock et al. (2007, Study 5) that the performance of women on a verbal task was lower when it followed a mathematics task explicitly primed to increase stereotype threat than when it followed the same task without such priming. Walker and Bridgeman compared the performance on a subsequent Critical Reading section for those who took the Mathematics section first with those who took the Critical Reading or Writing section first. Neither women's nor men's Critical Reading mean scores were lower when this section followed the Mathematics section than when it followed the other sections.

Stricker (2012) investigated changes in Black test takers' performance on the GRE General Test associated with Obama's 2008 presidential campaign. This study was modeled after one by Marx et al. (2009). In a field study motivated by the role-model effect in the Marx and Roman (2002) experiment—a competent woman experimenter enhanced women's test performance—Marx et al. observed that Black-White mean differences on a verbal ability test were reduced to nonsignificance at two points when Obama achieved concrete successes (after his nomination and after his election), though the differences were appreciable at other points. Stricker, using archival data for the GRE General Test's Verbal section, found that substantial Black-White differences persisted throughout the campaign and were virtually identical to the differences the year before the campaign.

The only ETS laboratory experiment thus far, by Lawrence Stricker and Isaac Bejar (2004), was a close replication of one by Spencer et al. (1999, Study 1). Spencer et al. found that women and men did not differ in their performance on an easy quantitative test, but they did differ on a hard one, consistent with the theoretical notion that stereotype threat is maximal when the test is difficult, at the limit of the test taker's ability. Stricker and Bejar used computer-adaptive versions of the GRE General Test, a standard version and one modified to produce a test that was easier but had comparable scores. Women's mean Quantitative scores, as well as their mean Verbal scores, did not differ on the easy and standard tests, and neither did the mean scores of the other participants: men, Blacks, and Whites.

In short, the ETS research to date has failed to find evidence of stereotype threat on operational tests in high-stakes settings, in common with work done elsewhere (Cullen et al. 2004, 2006). One explanation offered for this divergence from the results in other research studies is that motivation to perform well is heightened in a high-stakes setting, overriding any harmful effects of stereotype threat that might otherwise be found in the laboratory (Stricker and Ward 2004). The findings also suggest that changes in the test administration procedures or in the difficulty of the tests themselves are unlikely to ameliorate stereotype threat. In view of the limitations of field studies, the weight of laboratory evidence that document its robustness and potency, and its potential consequences for test validity (Stricker 2008), stereotype threat is a continuing concern at ETS.

## 13.6 Motivation

Motivation is at the center of psychological research, and its consequences for performance on tests, in school, and in other venues has been a long-standing subject for ETS investigations. Most of this research has focused on three related constructs: level of aspiration, need for achievement, and test anxiety. Level of aspiration, extensively studied by psychologists in the 1940s (e.g., see reviews by Lefcourt 1982; Powers 1986; Phares 1976), concerns the manner in which a person sets goals relative to that person's ability and past experience. Need for achievement, a very popular area of psychological research in the 1950s and 1960s (e.g., Atkinson 1957; McClelland et al. 1953), posits two kinds of motives in achievement-related situations: a motive to achieve success and a motive to avoid failure. Test anxiety is a manifestation of the latter. Research on test anxiety that focuses on its consequences for test performance has been a separate and active area of inquiry in psychology since the 1950s (e.g., see reviews by Spielberger and Vagg 1995; Zeidner 1998).

### 13.6.1 *Test Anxiety and Test Performance*

Several ETS studies have investigated the link between test anxiety and performance on ability and achievement tests. Two major studies by Donald Powers found moderate negative correlations between a test-anxiety measure and scores on the GRE General Test. In the first study (Powers 1986, 1988), when the independent contributions of the anxiety measure's Worry and Emotionality subscales were evaluated, only the Worry subscale was appreciably related to the test scores, suggesting that worrisome thoughts rather than physiological arousal affects test performance. The incidence of test anxiety was also reported. For example, 35% of test takers reported that they were tense and 36% that thoughts of doing poorly interfered with concentration on the test.

In the second study (Powers 2001), a comparison of the original, paper-based test and a newly introduced computer-adaptive version, a test-anxiety measure correlated similarly with the scores for the two versions. Furthermore, the mean level of test anxiety was slightly higher for the original version. These results indicate that the closer match between test-takers' ability and item difficulty provided by the computer-adaptive version did not markedly reduce test anxiety.

An ingenious experiment by French (1962) was designed to clarify the causal relationship between test anxiety and test performance. He manipulated test anxiety by administering sections of the SAT a few days before or after students took both the operational test and equivalent forms of these sections, telling the students that the results for the before and after sections would not be reported to colleges. The mean scores on these sections, which should *not* provoke test anxiety, were similar to those for sections administered with the SAT, which should provoke test anxiety,

after adjusting for practice effects. The before and after sections and the sections administered with the SAT correlated similarly with high school grades. The results in toto suggest that test anxiety did not affect performance on the test or change what it measured.

Connections between test anxiety and other aspects of test-taking behavior have been uncovered in studies not principally concerned with test anxiety. Stricker and Bejar (2004), using standard and easy versions of a computer-adaptive GRE General Test in a laboratory experiment, found that the mean level for a test-anxiety measure was lower for the easy version. This effect interacted with ethnicity (but not gender): White participants were affected but Black participants were not.

Lawrence Stricker and Gita Wilder (2002) reported small positive correlations between a test anxiety measure and the extent of preparation for the Pre-Professional Skills Tests (tests of academic skills used for admission to teacher education programs and for teacher licensing).

Finally, Stricker et al. (2004) observed minimal or small negative correlations between a test-anxiety measure and attitudes about the *TOEFL*<sup>®</sup> test and about admissions tests in general in a survey of TOEFL test takers in three countries.

### ***13.6.2 Test Anxiety/Defensiveness and Risk Taking and Creativity***

Several ETS studies documented the relation between test anxiety, usually in combination with defensiveness, and both risk taking and creativity. Nathan Kogan and Michael Wallach (1967b), Kogan's long-time collaborator at Duke University, investigated this relation in the context of the risky-shift phenomenon (i.e., group discussion enhances the risk-taking level of the group relative to the members' initial level of risk taking; Kogan and Wallach 1967a). In their study, small groups were formed on the basis of participants' scores on test-anxiety and defensiveness measures. Risk taking was measured by responses to hypothetical life situations. The risky-shift effect was greater for the pure test-anxious groups (high on test anxiety, low on defensiveness) than for the pure defensiveness groups (high on defensiveness, low on test anxiety). This outcome was consistent with the hypothesis that test anxious groups, fearful of failure, diffuse responsibility to reduce the possibility of personal failure, and defensiveness groups, being guarded, interact insufficiently for the risky-shift to occur.

Henry Alker (1969) found that a composite measure of test anxiety and defensiveness correlated substantially with a risk-taking measure (based on performance on SAT Verbal items)—those with low anxiety and low defensiveness took greater risks. In contrast, a composite of the McClelland standard Thematic Apperception Test (TAT) measure of need for achievement and a test-anxiety measure correlated only moderately with the same risk-taking measure—those with high need for achievement and low anxiety took more risks. This finding suggested that the Kogan and Wallach (1964, 1967a) theoretical formulation of the determinants of risk tak-

ing (based on test anxiety and defensiveness) was superior to the Atkinson-McClelland (Atkinson 1957; McClelland et al. 1953) formulation (based on need for achievement and test anxiety).

Wallach and Kogan (1965) observed a sex difference in the relationships of test anxiety and defensiveness measures with creativity (indexed by a composite of several measures). For boys, defensiveness was related to creativity but test anxiety was not—the more defensive were less creative; for girls, neither variable was related to creativity. For both boys and girls, the pure defensiveness subgroup (high defensiveness and low test anxiety) were the least creative, consistent with the idea that defensive people's cognitive performance is impaired in unfamiliar or ambiguous contexts.

Stephen Klein et al. (1969), as part of a larger experiment, reported an unanticipated curvilinear, U-shaped relationship between a test-anxiety measure and two creativity measures: Participants in the midrange of test anxiety had the lowest creativity scores. Klein et al. speculated that the low anxious participants make many creative responses because they do not fear ridicule for the poor quality of their responses; the high anxious participants make many responses, even though the quality is poor, because they fear a low score on the test; and the middling anxious participants make few responses because their two fears cancel each other out.

### ***13.6.3 Level of Aspiration or Need for Achievement and Academic Performance***

Another stream of ETS research investigated the connection between level of aspiration and need for achievement on the one hand, and performance in academic and other settings on the other. The results were mixed. Schultz and Ricciuti (1954) found that level of aspiration measures, based on a general ability test, a code learning task, and regular course examinations, did not correlate with college grades.

A subsequent study by John Hills (1958) used a questionnaire measure of level of aspiration in several areas, TAT measures of need for achievement in the same areas, and McClelland's standard TAT measure of need for achievement to predict law-school criteria. The level of aspiration and need for achievement measures did not correlate with grades or social activities in law school, but one or more of the level of aspiration measures had small or moderately positive correlations with undergraduate social activities and law-school faculty ratings of professional promise.

A later investigation by Albert Myers (1965) reported that a questionnaire measure of achievement motivation had a substantial positive correlation with high school grades.

### 13.6.4 Overview

Currently, research on motivation outside of the testing arena is not an active area of inquiry at ETS, but work on test anxiety and test performance continues, particularly when new kinds of tests and delivery systems for them are introduced. The investigations of the connection between test anxiety and both risk taking and creativity, and the work on test anxiety on operational tests, are significant contributions to knowledge in this field.

## 13.7 Conclusion

The scope of the research conducted by ETS that is covered in this chapter is extraordinary. The topics range across cognitive, personality, and social psychology. The methods include not only correlational studies, but also laboratory and field experiments, interviews, and surveys. And the populations studied are children, adults, psychiatric patients, and the general public, as well as students.

The work represents basic research in psychology, sometimes far removed from either education or testing, much less the development of products. Prosocial behavior is a case in point.

The research on almost all of the topics discussed has had major impacts on the field of psychology, even the short-lived work on prosocial behavior. Although the effects of some of the newer work, such as that on emotional intelligence, are too recent to gauge, as this chapter shows, that work continues a long tradition of contributions to these three fields of psychology.

**Acknowledgments** Thanks are due to Rachel Adler and Jason Wagner for their assistance in retrieving reports, articles, and archival material, and to Randy Bennett, Jeremy Burrus, and Donald Powers for reviewing a draft of this chapter.

## References

- Alker, H. A. (1969). Rationality and achievement: A comparison of the Atkinson-McClelland and Kogan-Wallach formulations. *Journal of Personality*, 37, 207–224. <https://doi.org/10.1111/j.1467-6494.1969.tb01741.x>
- American Educational Research Association, American Psychological Association, & National Council on Measurement in Education. (1999). *Standards for educational and psychological testing*. Washington, DC: American Educational Research Association.
- Atkinson, J. W. (1957). Motivational determinants of risk-taking behavior. *Psychological Review*, 64, 359–372. <https://doi.org/10.1037/h0043445>
- Austin, E. J. (2005). Emotional intelligence and emotional information processing. *Personality and Individual Differences*, 19, 403–414. <https://doi.org/10.1016/j.paid.2005.01.017>

- Beilock, S. L., Rydell, R. J., & McConnell, A. R. (2007). Stereotype threat and working memory: Mechanisms, alleviation, and spillover. *Journal of Experimental Psychology: General, 136*, 256–276. <https://doi.org/10.1037/0096-3445.136.2.256>
- Block, J. (1965). *The challenge of response sets—Unconfounding meaning, acquiescence, and social desirability in the MMPI*. New York: Appleton-Century-Crofts.
- Bryan, J. H., & Davenport, M. (1968). *Donations to the needy: Correlates of financial contributions to the destitute* (Research Bulletin No. RB-68-01). Princeton: Educational Testing Service. <http://dx.doi.org/10.1002/j.2333-8504.1968.tb00152.x>
- Bryan, J. H., & Test, M. A. (1967). Models and helping: Naturalistic studies in aiding behavior. *Journal of Personality and Social Psychology, 6*, 400–407. <https://doi.org/10.1037/h0024826>
- Bryan, J. H., & Walbek, N. H. (1970). Preaching and practicing generosity: Children's actions and reactions. *Child Development, 41*, 329–353. <https://doi.org/10.2307/1127035>
- Burton, L. J., & Fogarty, G. J. (2003). The factor structure of visual imagery and spatial abilities. *Intelligence, 31*, 289–318. [https://doi.org/10.1016/S0160-2896\(02\)00139-3](https://doi.org/10.1016/S0160-2896(02)00139-3)
- Carroll, J. B. (1993). *Human cognitive abilities: A survey of factor-analytic studies*. New York: Cambridge University Press. <https://doi.org/10.1017/CBO9780511571312>
- Cattell, R. B., & Gruen, W. (1955). The primary personality factors in 11-year-old children, by objective tests. *Journal of Personality, 23*, 460–478. <https://doi.org/10.1111/j.1467-6494.1955.tb01169.x>
- Cattell, R. B., & Scheier, I. H. (1959). Extension of meaning of objective test personality factors: Especially into anxiety, neuroticism, questionnaire, and physical factors. *Journal of General Psychology, 61*, 287–315. <https://doi.org/10.1080/00221309.1959.9710264>
- Cattell, R. B., Dubin, S. S., & Saunders, D. R. (1954). Verification of hypothesized factors in one hundred and fifteen objective personality test designs. *Psychometrika, 19*, 209–230. <https://doi.org/10.1007/BF02289186>
- Chapin, F. S. (1942). Preliminary standardization of a social insight scale. *American Sociological Review, 7*, 214–228. <https://doi.org/10.2307/2085176>
- Chauncey, H. (1947, July 13). [Notebook entry]. Henry Chauncey papers (Folder 1067). Carl C. Brigham library, Educational Testing Service, Princeton, NJ.
- Cliff, N. (1962). *Successful judgment in an interpersonal sensitivity task* (Research Bulletin No. RB-62-18). Princeton: Educational Testing Service. <http://dx.doi.org/10.1002/j.2333-8504.1962.tb00296.x>
- Cline, V. B. (1964). Interpersonal perception. *Progress in Experimental Personality Research, 2*, 221–284.
- Cronbach, L. J. (1946). Response sets and test validity. *Educational and Psychological Measurement, 6*, 475–494.
- Cronbach, L. J. (1950). Further evidence of response sets and test design. *Educational and Psychological Measurement, 10*, 3–31. <https://doi.org/10.1177/001316445001000101>
- Cronbach, L. J. (1990). *Essentials of psychological testing* (5th ed.). New York: Harper & Row.
- Cullen, M. J., Hardison, C. M., & Sackett, P. R. (2004). Using SAT-grade and ability-job performance relationships to test predictions derived from stereotype threat theory. *Journal of Applied Psychology, 89*, 220–230. <https://doi.org/10.1037/0021-9010.89.2.220>
- Cullen, M. J., Waters, S. D., & Sackett, P. R. (2006). Testing stereotype threat theory predictions for math-identified and non-math-identified students by gender. *Human Performance, 19*, 421–440. <https://doi.org/10.1037/0021-9010.89.2.220>
- Damarin, F., & Messick, S. (1965). *Response styles and personality variables: A theoretical integration of multivariate research* (Research Bulletin No. RB-65-10). Princeton: Educational Testing Service. <http://dx.doi.org/10.1002/j.2333-8504.1965.tb00967.x>
- Danaher, K., & Crandall, C. S. (2008). Stereotype threat in applied settings re-examined. *Journal of Applied Social Psychology, 38*, 1639–1655. <https://doi.org/10.1111/j.1559-1816.2008.00362.x>
- Edwards, A. L. (1957). *The social desirability variable in personality assessment and research*. New York: Dryden.



- Ekman, P., & Friesen, W. V. (1978). *Facial action coding system: A technique for the measurement of facial movement*. Palo Alto: Consulting Psychologists Press.
- Ekstrom, R. B., French, J. W., & Harman, H. H. (with Dermen, D.). (1976). *Manual for Kit of Factor-Referenced Cognitive Tests, 1976*. Princeton: Educational Testing Service.
- Ekstrom, R. B., French, J. W., & Harman, H. H. (1979). Cognitive factors: Their identification and replication. *Multivariate Behavioral Research Monographs*, No. 79-2.
- Frederiksen, N., & Schrader, W. (1951). *Adjustment to college—A study of 10,000 veteran and non-veteran students in sixteen American colleges*. Princeton: Educational Testing Service.
- French, J. W. (1951a). *Conference on factorial studies of aptitude and personality measures* (Research Memorandum No. RM-51-20). Princeton: Educational Testing Service.
- French, J. W. (1951b). The description of aptitude and achievement factors in terms of rotated factors. *Psychometric Monographs*, No. 5.
- French, J. W. (1952). *Validity of group Rorschach and Rosenzweig P. F. Study for adaptability at the U.S. coast guard academy* (Research Memorandum No. RM-52-02). Princeton: Educational Testing Service.
- French, J. W. (1953). *Selected tests for reference factors—Draft of final report* (Research Memorandum No. RM-53-04). Princeton: Educational Testing Service.
- French, J. W. (1954). *Manual for Kit of Selected Tests for Reference Aptitude and Achievement Factors*. Princeton: Educational Testing Service.
- French, J. W. (1958). *Working plans for the reference test project* (Research Memorandum No. RM-58-10). Princeton: Educational Testing Service.
- French, J. W. (1962). Effect of anxiety on verbal and mathematical examination scores. *Educational and Psychological Measurement*, 22, 553–564. <https://doi.org/10.1177/001316446202200313>
- French, J. W., Ekstrom, R. B., & Price, L. A. (1963). *Manual for Kit of Reference Tests for Cognitive Factors*. Princeton: Educational Testing Service.
- Goleman, D. (1995). *Emotional intelligence: Why it can matter more than IQ*. New York: Bantam.
- Greenwald, A. G., McGhee, D. E., & Schwartz, J. L. K. (1998). Measuring individual differences in implicit cognition: The implicit association test. *Journal of Personality and Social Psychology*, 74, 1464–1480. <https://doi.org/10.1037/0022-3514.74.6.1464>
- Harman, H. H. (1975). *Final report of research on assessing human abilities* (Project Report No. PR-75-20). Princeton: Educational Testing Service.
- Hegarty, M., Shah, P., & Miyake, A. (2000). Constraints on using the dual-task methodology to specify the degree of central executive involvement in cognitive tasks. *Memory & Cognition*, 28, 373–385. <https://doi.org/10.3758/BF03198553>
- Hendricks, M., Guilford, J. P., & Hoepfner, R. (1969). *Measuring creative social intelligence* (Psychological Laboratory Report No. 42). Los Angeles: University of Southern California.
- Hills, J. R. (1958). Needs for achievement, aspirations, and college criteria. *Journal of Educational Psychology*, 49, 156–161. <https://doi.org/10.1037/h0047283>
- Hofstee, W. K. B., ten Berge, J. M. F., & Hendriks, A. A. J. (1998). How to score questionnaires. *Personality and Individual Differences*, 25, 897–909. [https://doi.org/10.1016/S0191-8869\(98\)00086-5](https://doi.org/10.1016/S0191-8869(98)00086-5)
- Hsee, C. K., & Hastie, R. (2006). Decision and experience: Why don't we choose what makes us happy? *Trends in Cognitive Science*, 10, 31–37. <https://doi.org/10.1016/j.tics.2005.11.007>
- Jackson, D. N. (1967). Acquiescence response styles: Problems of identification and control. In I. A. Berg (Ed.), *Response set in personality assessment* (pp. 71–114). Chicago: Aldine.
- Jackson, D. N., & Messick, S. (1958). Content and style in personality assessment. *Psychological Bulletin*, 55, 243–252. <https://doi.org/10.1037/h0045996>
- Jackson, D. N., & Messick, S. (1961). Acquiescence and desirability as response determinants on the MMPI. *Educational and Psychological Measurement*, 21, 771–790. <https://doi.org/10.1177/001316446102100402>
- Jackson, D. N., & Messick, S. (1962a). Response styles and the assessment of psychopathology. In S. Messick & J. Ross (Eds.), *Measurement in personality and cognition* (pp. 129–155). New York: Wiley.

- Jackson, D. N., & Messick, S. (1962b). Response styles on the MMPI: Comparison of clinical and normal samples. *Journal of Abnormal and Social Psychology*, *65*, 285–299. <https://doi.org/10.1037/h0045340>
- Klein, S. P., Frederiksen, N., & Evans, F. R. (1969). Anxiety and learning to formulate hypotheses. *Journal of Educational Psychology*, *60*, 465–475. <https://doi.org/10.1037/h0028351>
- Kogan, N., & Wallach, M. A. (1964). *Risk taking: A study in cognition and personality*. New York: Holt, Rinehart and Winston.
- Kogan, N., & Wallach, M. A. (1967a). Effects of physical separation of group members upon group risk taking. *Human Relations*, *20*, 41–49. <https://doi.org/10.1177/001872676702000104>
- Kogan, N., & Wallach, M. A. (1967b). Group risk taking as a function of members' anxiety and defensiveness levels. *Journal of Personality*, *35*, 50–63. <https://doi.org/10.1111/j.1467-6494.1967.tb01415.x>
- Krebs, D. L., & Miller, D. T. (1985). Altruism and aggression. In G. Lindzey & E. Aronson (Eds.), *Handbook of social psychology: Vol. 2. Special fields and applications* (3rd ed., pp. 1–71). New York: Random House.
- Latané, B., & Darley, J. M. (1970). *The unresponsive bystander: Why doesn't he help?* New York: Appleton-Century-Crofts.
- Lefcourt, H. M. (1982). *Locus of control: Current trends in theory and research* (2nd ed.). Hillsdale: Erlbaum.
- MacCann, C., & Roberts, R. D. (2008). New paradigms for assessing emotional intelligence: Theory and data. *Emotions*, *8*, 540–551. <https://doi.org/10.1037/a0012746>
- MacCann, C., Schulze, R., Matthews, G., Zeidner, M., & Roberts, R. D. (2008). Emotional intelligence as pop science, misled science, and sound science: A review and critical synthesis of perspectives from the field of psychology. In N. C. Karafyllis & G. Ulshofer (Eds.), *Sexualized brains—Scientific modeling of emotional intelligence from a cultural perspective* (pp. 131–148). Cambridge: MIT Press.
- MacCann, C., Wang, L., Matthews, G., & Roberts, R. D. (2010). Emotional intelligence and the eye of the beholder: Comparing self- and parent-rated situational judgments in adolescents. *Journal of Research in Personality*, *44*, 673–676. <https://doi.org/10.1016/j.jrp.2010.08.009>
- MacCann, C., Fogarty, G. J., Zeidner, M., & Roberts, R. D. (2011). Coping mediates the relationship between emotional intelligence (EI) and academic achievement. *Contemporary Educational Psychology*, *36*, 60–70. <https://doi.org/10.1016/j.cedpsych.2010.11.002>
- Manning, R., Levine, M., & Collins, A. (2007). The Kitty Genovese murder and the social psychology of helping behavior—The parable of the 38 witnesses. *American Psychologist*, *68*, 555–562. <https://doi.org/10.1037/0003-066X.62.6.555>
- Marx, D. M., & Roman, J. S. (2002). Female role models: Protecting women's math test performance. *Personality and Social Psychology Bulletin*, *28*, 1183–1193.
- Marx, D. M., Ko, S. J., & Friedman, R. A. (2009). The “Obama effect”: How a salient role model reduces race-based differences. *Journal of Experimental Social Psychology*, *45*, 953–956. <https://doi.org/10.1016/j.jesp.2009.03.012>
- Mayer, J. D., & Salovey, P. (1997). What is emotional intelligence? In P. Salovey & D. Sluyter (Eds.), *Emotional development and emotional intelligence: Implications for educators* (pp. 3–31). New York: Basic Books.
- Mayer, J. D., Salovey, P., & Caruso, D. R. (2002). *Mayer-Salovey-Caruso Emotional Intelligence Test (MSCEIT) user's manual*. Toronto: Multi-Health Systems.
- Mayer, J. D., Roberts, R. D., & Barsade, S. G. (2008). Human abilities: Emotional intelligence. *Annual Review of Psychology*, *59*, 507–536. <https://doi.org/10.1146/annurev.psych.59.103006.093646>
- McClelland, D. C., Atkinson, J. W., Clark, R. A., & Lowell, E. L. (1953). *The achievement motive*. New York: Appleton-Century-Crofts. <https://doi.org/10.1037/11144-000>

- McCrae, R. R., Herbst, J., & Costa Jr., P. T. (2001). Effects of acquiescence on personality factor structures. In R. Riemann, F. M. Spinath, & F. R. Ostendorf (Eds.), *Personality and temperament: Genetics, evolution, and structure* (pp. 216–231). Langerich: Pabst.
- McGrath, R. E., Mitchell, M., Kim, B. H., & Hough, L. (2010). Evidence for response bias as a source of error variance in applied assessment. *Psychological Bulletin*, *136*, 450–470. <https://doi.org/10.1037/a0019216>
- Melville, S. D., & Frederiksen, N. (1952). Achievement of freshman engineering students and the Strong Vocational Interest Blank. *Journal of Applied Psychology*, *36*, 169–173. <https://doi.org/10.1037/h0059101>
- Messick, S. (1967). The psychology of acquiescence: An interpretation of research evidence. In I. A. Berg (Ed.), *Response set in personality assessment* (pp. 115–145). Chicago, IL: Aldine.
- Messick, S. (1991). Psychology and methodology of response styles. In R. E. Snow & D. E. Wiley (Eds.), *Improving inquiry in social science—A volume in honor of Lee J. Cronbach* (pp. 161–200). Hillsdale: Erlbaum.
- Midlarsky, E., & Bryan, J. H. (1967). Training charity in children. *Journal of Personality and Social Psychology*, *5*, 408–415. <https://doi.org/10.1037/h0024399>
- Myers, A. E. (1965). Risk taking and academic success and their relation to an objective measure of achievement motivation. *Educational and Psychological Measurement*, *25*, 355–363. <https://doi.org/10.1177/001316446502500206>
- Nogee, P. (1950). *A preliminary study of the "Social Situations Test"* (Research Memorandum No. RM-50-22). Princeton: Educational Testing Service.
- O'Sullivan, M., & Guilford, J. P. (1975). Six factors of behavioral cognition: Understanding other people. *Journal of Educational Measurement*, *12*, 255–271. <https://doi.org/10.1111/j.1745-3984.1975.tb01027.x>
- Ones, D. S., Viswesvaran, C., & Reiss, A. D. (1996). Role of social desirability in personality testing for personnel selection: The red herring. *Journal of Applied Psychology*, *81*, 660–679. <https://doi.org/10.1037/0021-9010.81.6.660>
- Orchard, B., MacCann, C., Schulze, R., Matthews, G., Zeidner, M., & Roberts, R. D. (2009). New directions and alternative approaches to the measurement of emotional intelligence. In C. Stough, D. H. Saklofske, & J. D. A. Parker (Eds.), *Assessing human intelligence—Theory, research, and applications* (pp. 321–344). New York: Springer. [https://doi.org/10.1007/978-0-387-88370-0\\_17](https://doi.org/10.1007/978-0-387-88370-0_17)
- Paulhus, D. L. (1984). Two-component models of socially desirable responding. *Journal of Personality and Social Psychology*, *46*, 598–609. <https://doi.org/10.1037/0022-3514.46.3.598>
- Paulhus, D. L. (2002). Socially desirable responding: The evolution of a construct. In H. I. Braun, D. N. Jackson, & D. E. Wiley (Eds.), *The role of constructs in psychological and educational measurement* (pp. 49–69). Mahwah: Erlbaum.
- Paulhus, D. L., & Vazire, S. (2007). The self-report method. In R. W. Robins, R. C. Fraley, & R. F. Krueger (Eds.), *Handbook of research methods in personality psychology* (pp. 224–239). New York: Guilford.
- Penner, L. A., Dovidio, J. F., Pillavin, J. A., & Schroeder, D. A. (2005). Prosocial behavior: Multilevel perspectives. *Annual Review of Psychology*, *56*, 365–392. <https://doi.org/10.1146/annurev.psych.56.091103.070141>
- Phares, E. J. (1976). *Locus of control in personality*. Morristown: General Learning Press.
- Powers, D. E. (1986). *Test anxiety and the GRE General Test* (GRE Board Professional Report No. 83-17P). Princeton: Educational Testing Service. <http://dx.doi.org/10.1002/j.2330-8516.1986.tb00200.x>
- Powers, D. E. (1988). Incidence, correlates, and possible causes of test anxiety in graduate admissions testing. *Advances in Personality Assessment*, *7*, 49–75.
- Powers, D. E. (2001). Test anxiety and test performance: Comparing paper-based and computer-adaptive versions of the Graduate Record Examinations (GRE) General Test. *Journal of Educational Computing Research*, *24*, 249–273. <https://doi.org/10.2190/680W-66CR-QRP7-CL1F>

- Ricciuti, H. N. (1951). *A comparison of leadership ratings made and received by student raters* (Research Memorandum No. RM-51-04). Princeton: Educational Testing Service.
- Roberts, R. D., Schulze, R., Zeidner, M., & Matthews, G. (2005). Understanding, measuring, and applying emotional intelligence: What have we learned? What have we missed? In R. Schulze & R. D. Roberts (Eds.), *Emotional intelligence—An international handbook* (pp. 311–341). Gottingen: Hogrefe & Huber.
- Roberts, R. D., Schulze, R., O'Brien, K., McCann, C., Reid, J., & Maul, A. (2006). Exploring the validity of the Mayer-Salovey-Caruso Emotional Intelligence Test (MSCEIT) with established emotions measures. *Emotions*, 6, 663–669.
- Roberts, R. D., Schulze, R., & MacCann, C. (2008). The measurement of emotional intelligence: A decade of progress? In G. Boyle, G. Matthews, & D. Saklofske (Eds.), *The Sage handbook of personality theory and assessment: Vol 2. Personality measurement and testing* (pp. 461–482).
- Roberts, R. D., MacCann, C., Matthews, G., & Zeidner, M. (2010). Emotional intelligence: Toward a consensus of models and measures. *Social and Personality Psychology Compass*, 4, 821–840. <https://doi.org/10.1111/j.1751-9004.2010.00277.x>
- Rorer, L. G., & Goldberg, L. R. (1965a). Acquiescence and the vanishing variance component. *Journal of Applied Psychology*, 49, 422–430. <https://doi.org/10.1037/h0022754>
- Rorer, L. G., & Goldberg, L. R. (1965b). Acquiescence in the MMPI? *Educational and Psychological Measurement*, 25, 801–817. <https://doi.org/10.1177/001316446502500311>
- Rosenhan, D. (1969). Some origins of concern for others. In P. H. Mussen, J. Langer, & M. V. Covington (Eds.), *Trends and issues in developmental psychology* (pp. 134–153). New York: Holt, Rinehart and Winston.
- Rosenhan, D. (1970). The natural socialization of altruistic autonomy. In J. Macaulay & L. Berkowitz (Eds.), *Altruism and helping behavior—Social psychological studies of some antecedents and consequences* (pp. 251–268). New York: Academic Press.
- Rosenhan, D. L. (1972). Learning theory and prosocial behavior. *Journal of Social Issues*, 28(3), 151–163. <https://doi.org/10.1111/j.1540-4560.1972.tb00037.x>
- Rosenhan, D., & White, G. M. (1967). Observation and rehearsal as determinants of prosocial behavior. *Journal of Personality and Social Psychology*, 5, 424–431. <https://doi.org/10.1037/h0024395>
- Rosenthal, R., Hall, J. A., DiMatteo, M. R., Rogers, P. L., & Archer, D. (1979). *Sensitivity to non-verbal communication—The PONS test*. Baltimore: Johns Hopkins University Press. <https://doi.org/10.1016/b978-0-12-761350-5.50012-4>
- Salovey, P., & Mayer, J. D. (1990). Emotional intelligence. *Imagination, Cognition, and Personality*, 9, 185–211. <https://doi.org/10.2190/DUGG-P24E-52WK-6CDG>
- Saucier, G., & Goldberg, L. R. (2002). Assessing the big five: Applications of 10 psychometric criteria to the development of marker scales. In B. de Raad & M. Perugini (Eds.), *Big five assessment* (pp. 29–58). Gottingen: Hogrefe & Huber.
- Schultz, D. G., & Ricciuti, H. N. (1954). Level of aspiration measures and college achievement. *Journal of General Psychology*, 51, 267–275. <https://doi.org/10.1080/00221309.1954.9920226>
- Schulze, R., Wilhelm, O., & Kyllonen, P. C. (2007). Approaches to the assessment of emotional intelligence. In G. Matthews, M. Zeidner, & R. D. Roberts (Eds.), *The science of emotional intelligence—Knowns and unknowns* (pp. 199–229). New York: Oxford University Press.
- Spearman, C. (1904). “General intelligence” objectively determined and measured. *American Journal of Psychology*, 15, 201–293. <https://doi.org/10.2307/1412107>
- Spencer, S. J., Steele, C. M., & Quinn, D. M. (1999). Stereotype threat and women’s math performance. *Journal of Experimental Social Psychology*, 35, 4–28. <https://doi.org/10.1006/jesp.1998.1373>

- Spielberger, C. D., & Vagg, P. R. (Eds.). (1995). *Test anxiety: Theory, assessment, and treatment*. Washington, DC: Taylor & Francis.
- Steele, C. M. (1997). A threat in the air: How stereotypes shape intellectual identity and performance. *American Psychologist*, 52, 613–629. <https://doi.org/10.1037/0003-066X.52.6.613>
- Steele, C. M., & Aronson, J. (1995). Stereotype threat and the intellectual test performance of African Americans. *Journal of Personality and Social Psychology*, 69, 797–811. <https://doi.org/10.1037/0022-3514.69.5.797>
- Stricker, L. J. (2008). *The challenge of stereotype threat for the testing community* (Research Memorandum No. RM-08-12). Princeton: Educational Testing Service.
- Stricker, L. J. (2012). *Testing: It's not just psychometrics* (Research Memorandum No. RM-12-07). Princeton: Educational Testing Service.
- Stricker, L. J., & Bejar, I. (2004). Test difficulty and stereotype threat on the GRE General Test. *Journal of Applied Social Psychology*, 34, 563–597. <https://doi.org/10.1111/j.1559-1816.2004.tb02561.x>
- Stricker, L. J., & Rock, D. A. (1990). Interpersonal competence, social intelligence, and general ability. *Personality and Individual Differences*, 11, 833–839. [https://doi.org/10.1016/0191-8869\(90\)90193-U](https://doi.org/10.1016/0191-8869(90)90193-U)
- Stricker, L. J., & Ward, W. C. (2004). Stereotype threat, inquiring about test takers' ethnicity and gender, and standardized test performance. *Journal of Applied Social Psychology*, 34, 665–693. <https://doi.org/10.1111/j.1559-1816.2004.tb02564.x>
- Stricker, L. J., & Ward, W. C. (2008). Stereotype threat in applied settings re-examined: A reply. *Journal of Applied Social Psychology*, 38, 1656–1663. <https://doi.org/10.1111/j.1559-1816.2008.00363.x>
- Stricker, L. J., & Wilder, G. Z. (2002). Why don't test takers prepare for the Pre-professional Skills Test? *Educational Assessment*, 8, 259–277. [https://doi.org/10.1207/S15326977EA0803\\_03](https://doi.org/10.1207/S15326977EA0803_03)
- Stricker, L. J., Wilder, G. Z., & Rock, D. A. (2004). Attitudes about the computer-based Test of English as a Foreign Language. *Computers in Human Behavior*, 20, 37–54. [https://doi.org/10.1016/S0747-5632\(03\)00046-3](https://doi.org/10.1016/S0747-5632(03)00046-3)
- Test, M. A., & Bryan, J. H. (1969). The effects of dependency, models, and reciprocity upon subsequent helping behavior. *Journal of Social Psychology*, 78, 205–212. <https://doi.org/10.1080/00224545.1969.9922357>
- Thorndike, E. L. (1920, January). Intelligence and its uses. *Harper's Monthly Magazine*, 140, 227–235.
- Thurstone, L. L. (1935). *The vectors of mind—Multiple-factor analysis for the isolation of primary traits*. Chicago: University of Chicago Press. <https://doi.org/10.1037/10018-000>
- Thurstone, L. L. (1938). Primary mental abilities. *Psychometric Monographs*, No. 1.
- Walker, M. E., & Bridgeman, B. (2008). *Stereotype threat spillover and SAT scores* (College Board Report No. 2008–2). New York: College Board.
- Wallach, M. A., & Kogan, N. (1965). *Modes of thinking in young children—A study of the creativity-intelligence distinction*. New York: Holt, Rinehart and Winston.
- Walters, A. M., Lee, S., & Trapani, C. (2004). *Stereotype threat, the test-center environment, and performance on the GRE General Test* (GRE Board Research Report No. 01-03R). Princeton: Educational Testing Service. <http://dx.doi.org/10.1002/j.2333-8504.2004.tb01964.x>
- White, G. M. (1972). Immediate and deferred effects of model observation and guided and unguided rehearsal on donating and stealing. *Journal of Personality and Social Psychology*, 21, 139–148. <https://doi.org/10.1037/h0032308>
- Wiggins, J. S. (1973). *Personality and prediction: Principles of personality assessment*. Reading: Addison-Wesley.

- Wispe, L. G. (1972). Positive forms of social behavior: An overview. *Journal of Social Issues*, 28(3), 1–19. <https://doi.org/10.1111/j.1540-4560.1972.tb00029.x>
- Zeidner, M. (1998). *Test anxiety: The state of the art*. New York: Plenum.

**Open Access** This chapter is licensed under the terms of the Creative Commons Attribution-NonCommercial 2.5 International License (<http://creativecommons.org/licenses/by-nc/2.5/>), which permits any noncommercial use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the chapter's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.

