propriate, 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*® 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

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
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

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 nonverbal 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

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

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

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.

Wispé, 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.

Chapter 14
Research on Cognitive, Personality, and Social Psychology: II

Nathan Kogan

This is the second of two chapters describing research at Educational Testing Service (ETS) on cognitive, personality, and social psychology since its founding in 1947. The first chapter, Chap. 13 by Lawrence Stricker, also appears in this volume. Topics in these fields were selected for attention because they were the focus of extensive and significant ETS research. This chapter covers these topics: in cognitive psychology, creativity; in personality psychology, cognitive styles and kinesthetic aftereffect; and in social psychology, risk taking.

14.1 Creativity

Research on creativity thrived at ETS during the 1960s and 1970s. Three distinct strands of work can be distinguished. One of these strands was based largely on studies of children, with an emphasis on performance in the domain of divergent-thinking abilities. A second strand involved the construction of measures of scientific thinking and utilized samples of young adults. A third strand featured an emphasis on the products of creativity, mainly using young adult samples. The three strands were not entirely independent of each other as some studies explored possible links between the divergent-thinking and scientific-thinking domains or between ratings of products and characteristics of the individuals who produced them.

Lawrence J. Stricker edited a final draft of this chapter that Nathan Kogan completed before his death.

N. Kogan (✉)
Educational Testing Service, Princeton, NJ, USA
e-mail: researchreports@ets.org

R.E. Bennett, M. von Davier (eds.), *Advancing Human Assessment*,
Methodology of Educational Measurement and Assessment,
DOI 10.1007/978-3-319-58689-2_14

14.1.1 Divergent Thinking

We begin with studies in the divergent-thinking domain that employed children ranging across the preschool to primary-school years. The volume published by ETS scientist Nathan Kogan and Michael Wallach, his longtime collaborator at Duke University (Wallach and Kogan 1965a), set the tone for much of the research that followed. A major goal of that investigation was to bring clarity to the discriminant validity issue—whether divergent-thinking abilities could be statistically separated from the convergent thinking required by traditional tests of intellectual ability. In a paper by Thorndike (1966), evidence for such discriminant validity in investigations by Guilford and Christensen (1956) and by Getzels and Jackson (1962) was found to be lacking. A similar failure was reported by Wallach (1970) for the Torrance (1962) divergent-thinking test battery—the correlations within the divergent-thinking battery were of approximately the same magnitude as these tests' correlations with a convergent-thinking test. Accordingly, the time was ripe for another effort at psychometric separation of the divergent- and convergent-thinking domains.

Wallach and Kogan (1965a) made two fundamental changes in the research paradigms that had been used previously. They chose to purify the divergent-thinking domain by employing only ideational-fluency tasks and presented these tasks as games, thereby departing from the mode of administration typical of convergent-thinking tests. The rationale for these changes can be readily spelled out. Creativity in the real world involves the generation of new ideas, and this is what ideational-fluency tests attempt to capture. Of course, the latter represents a simple analogue of the former, but the actual relationship between them rests on empirical evidence (which is mixed at the present time). The choice of a game-like atmosphere was intended to reduce the test anxiety from which numerous test takers suffer when confronted with typical convergent-thinking tests.

The major outcome of these two modifications was the demonstration of both convergent validity of measures of the divergent- and convergent-thinking domains and the discrimination between them as reflected in near-zero correlations in a sample of fifth-grade children. As evidence has accumulated since the Wallach and Kogan study (Wallach and Kogan 1965a), the trend is toward a low positive correlation between measures of the two domains. For example, Silvia (2008), employing latent variable analysis, reanalyzed the Wallach and Kogan data and reported a significant correlation of .20, consistent with the predominant outcome of the majority of studies directed to the issue. It may well be a pseudo-issue at this point in time, reflecting the selectivity of the sample employed. As the range of IQ in a sample declines, its correlation with divergent-thinking measures should obviously decline as well. Thus, one would not expect to find divergent- and convergent-thinking tests correlated in a sample selected for giftedness.

The ideational-fluency tests developed by Wallach and Kogan (1965a) were scored for fluency and uniqueness. The two were highly correlated, consistent with what was expected by the principal theoretical conceptualization at the time—

Mednick's (1962) associative theory of creativity. In that theory, the associative process for divergent-thinking items initially favors common associates, and only with continued association would unique and original associations be likely to emerge. Accordingly, fluency represents the path through which originality was achieved. Individual differences in divergent-thinking performance are explained by the steepness-shallowness of the associative hierarchy. Low creatives exhibit a steep gradient in which strong common responses, upon their exhaustion, leave minimal response strength for the emergence of uncommon associates. High creatives, by contrast, demonstrate a shallow gradient in which response strength for common associates is weaker, allowing the person enough remaining power to begin emitting uncommon associates.

In a recent article, Silvia et al. (2008) ignored the Mednick (1962) formulation, criticized the scoring of divergent-thinking responses for uniqueness, concluded that scoring them for quality was psychometrically superior, and advocated that the administration of divergent-thinking tests urge test takers to be creative. A critical commentary on this work by Silvia and his associates appeared in the same issue of the journal (Kogan 2008). Particularly noteworthy is the indication that approximately 45 years after its publication, the issues raised in the Wallach and Kogan (1965a) volume remain in contention.

Beyond the topic of the creativity-intelligence (divergent vs. convergent thinking) distinction, the construct validity of divergent-thinking tests came under exploration. What psychological processes (beyond Mednick's 1962, response hierarchies) might account for individual differences in divergent-thinking performance? Pankove and Kogan (1968) suggested that tolerance for risk of error might contribute to superior divergent-thinking performance in elementary school children. A motor skill task (a shuffleboard game) allowed children to adjust their preferred risk levels by setting goal posts closer or further apart to make the task harder or easier, respectively. Children who challenged themselves by taking greater risks on the shuffleboard court (with motor skill statistically controlled) also generated higher scores on a divergent-thinking test, Alternate Uses.

In a provocative essay, Wallach (1971) offered the hypothesis that performance on divergent-thinking tests might be motivationally driven. In other words, test takers might vary in setting personal standards regarding an adequate number of responses. Some might stop well before their cognitive repertoire is exhausted, whereas others might continue to generate responses in a compulsive fashion. This hypothesis implies that the application of an incentive to continue for low-level responders should attenuate the range of fluency scores. Ward et al. (1972) tested this hypothesis in a sample of disadvantaged children by offering an incentive of a penny per response. The incentive increased the number of ideas relative to a control group but did not reduce the range of individual differences. Rather, the incentive added a constant to performance so that the original ordering of the children on the fluency dimension remained intact. In sum, the study bolstered the case for cognitive processes and repertoires underlying divergent-thinking performance and undermined the motivational claim that it is a simple matter of when one chooses to stop responding.

To designate divergent thinking as an indicator of creativity is credible only if divergent-thinking performance is predictive of a real-world criterion that expert judges would acknowledge to be relevant to creativity. This is the validity issue that has been examined in both its concurrent and long-term predictive forms. The concurrent validity of divergent-thinking performance has proven to be rather robust. Thus, third-grade and fourth-grade children's scores on the Wallach and Kogan (1965a) tasks correlated significantly with the originality and aesthetic quality of their art products, as evaluated by qualified judges (Wallbrown and Huelsman 1975). And college freshmen's scores on these tasks correlated significantly with their extracurricular attainments in leadership, art, writing, and sciences in their secondary-school years, whereas their *SAT*® scores did not (Wallach and Wing 1969). Efforts to predict future talented accomplishments from current divergent-thinking performance have yielded more equivocal outcomes. Kogan and Pankove (1972, 1974) failed to demonstrate predictive validity of fifth-grade Wallach and Kogan assessments against 10th-grade and 12th-grade accomplishments in extracurricular activities in the fields of art, writing, and science. On the other hand, Plucker's (1999) reanalysis of original data from the Torrance Tests of Creative Thinking (Torrance 1974) is suggestive of the predictive validity of that instrument.

The issue of predictive validity from childhood to adulthood continues to reverberate to the present day, with Kim (2011) insisting that the evidence is supportive for the Torrance tests while Baer (2011) notes that the adulthood creativity criterion employed is based exclusively on self-reports, hence rendering the claim for predictive validity highly suspect. Indeed, Baer extends his argument to the point of recommending that the Torrance creativity tests be abandoned.

Can the Mednick (1962) associative model of creativity be generalized to young children? Ward (1969b) offered an answer to this question by administering some of the Wallach and Kogan (1965a) tasks to seven- and eight-year old boys. The model was partially confirmed in the sense that the response rate (and the number of common responses) decreased over time while uniqueness increased over time. On the other hand, individual differences in divergent thinking did not seem to influence the steepness versus shallowness of the response gradients. Ward suggested that cognitive repertoires are not yet fully established in young children, while motivational factors (e.g., task persistence over time) that are not part of Mednick's theoretical model loom large.

Although Mednick's (1962) associative theory of creativity can explain individual differences in divergent-thinking performance, he chose to develop a creativity test—the Remote Associates Test (RAT; Mednick and Mednick 1962)—with a convergent-thinking structure. Items consist of verbal triads for which the test taker is required to find a word that is associatively linked to each of the three words in the triad. An example is "mouse, sharp, blue"; cheese is the answer. It is presumed that the correct answer to each item requires an associative verbal flow, with conceptual thinking of no value for problem solution. Working with children in the fourth-grade to sixth-grade, Ward (1975) administered the Wallach and Kogan (1965a) divergent-thinking tasks and alternate forms of the RAT, as well as IQ and achievement tests. Both forms of the RAT were substantially related to the IQ and

achievement measures (*r*'s ranging from .50 to .64). Correlations of the RAT with the Wallach and Kogan tasks ranged from nonsignificant to marginally significant (*r*'s ranging from .19 to .34). These results demonstrated that the associative process is not similar across divergent- and convergent-thinking creativity measures and that the latter's strong relation to IQ and achievement indicates the RAT "represents an unusual approach to the measurement of general intellectual ability" (Ward 1975, p. 94), rather than being a creativity measure.

Among the different explanations for variation in divergent-thinking performance, breadth of attention deployment (Wallach 1970) has been considered important. This process has both an internal and external component, with the former reflecting the adaptive scanning of personal cognitive repertoires and the latter indicative of adaptive scanning of one's immediate environment. A demonstration of the latter can be found in Ward's (1969a) investigation of nursery school children's responses in cue-rich and cue-poor environments. Recognition and application of such cues enhanced divergent-thinking performance, as the cues were directly relevant to the divergent-thinking items presented to the child. Some of the cues in the cue-rich environment were highly salient; some were more subtle; and for some items, no cues were offered. Children were classified as more or less creative based on their pre-experimental divergent-thinking performance. Comparison of high and low creative children revealed no divergent-thinking performance difference with salient cues, and significant performance superiority for high creatives with subtle cues. The low-creative children performed worse in the cue-rich environment than in the cue-poor environment, suggesting that the cue-rich environment was distracting for them. Hence, children who performed well on divergent-thinking items in standard cue-poor conditions by virtue of internal scanning also took advantage of environmental cues for divergent-thinking items by virtue of adaptive external scanning.

To conclude the present section on children's divergent thinking, consider the issue of strategies children employed in responding to divergent-thinking tasks under test-like and game-like conditions (Kogan and Morgan 1969). In a verbal alternate-uses task, children (fifth graders) generated higher levels of fluency and uniqueness in a test-like than in a game-like condition. Yet, a spontaneous-flexibility test (number of response categories) showed no difference between the task contexts. Kogan and Morgan (1969) argued that a test-like condition stimulated a category-exhaustion strategy. Thus, when asked to list alternative uses for a knife, children seized upon some pivotal activity (such as cutting) and proceeded to exhaust exemplars that flow from it (e.g., cutting bread, butter, fruit). A child might eventually think of something to cut that is unique to the sample. Such a strategy is obviously antithetical to enhanced spontaneous-flexibility. The test-game difference did not emerge for the Wallach and Kogan (1965a) figural pattern-meanings task. This outcome may be attributed to the inability of a category-exhaustion strategy to work with a figural task where almost every response is likely to be a category in its own right. In sum, verbal and figural divergent-thinking tasks might elicit distinctive cognitive strategies in children that are moderated by the task context. Further discussion of the issue can be found in Kogan (1983).

14.1.2 Scientific Thinking

In an Office of Naval Research technical report, ETS scientist Norman Frederiksen (1959) described the development of the Formulating Hypotheses test that asks test takers to assume the role of a research investigator attempting to account for a set of results presented in tabular or figural form. An example of the latter is a graph demonstrating that "rate of death from infectious diseases has decreased markedly from 1900, while rate of death from diseases of old age has increased." Examples of possible explanations for the findings orient the test taker to the type of reasoning required by the test. Eight items were constructed, and in its initial version scoring simply involved a count of the number of hypotheses advanced. Subsequently, as a pool of item responses became available, each response could be classified as acceptable or not. The number of acceptable responses generated could then be treated as a quality score.

Publications making use of this test began with an article by Klein et al. (1969). In that study of a college undergraduate sample, Klein et al. explored the influence of feedback after each item relative to a control group with no feedback. The number of hypotheses offered with feedback increased significantly relative to the number for the control group, but no experimental-control difference was found for acceptable (higher quality) hypotheses. Further, no experimental-control difference was observed for Guilford's (1967) Consequences test, a measure of divergent production, indicating no transfer effects. An anxiety scale was also administered with the expectation that anxiety would enhance self-censorship on the items, which in turn would be mitigated in the feedback treatment as anxious participants become aware of the vast array of hypotheses available to them. No such effect was obtained. Klein et al. also examined the possibility that intermediate levels of anxiety would be associated with maximal scores on the test consistent with the U-shaped function of motivational arousal and performance described by Spence and Spence (1966). Surprisingly, this hypothesis also failed to be confirmed. In sum, this initial study by Klein et al. demonstrated the potential viability of the Formulating Hypotheses test as a measure of scientific thinking despite its failure to yield anticipated correlates.

A further advance in research on this test is displayed in a subsequent study by Frederiksen and Evans (1974). As in the previous investigation, this one featured an experimental-control contrast, but two treatments were now employed. Participants (college undergraduates) were exposed to either quantity or quality models. In the former, the feedback following each item consisted of a lengthy list of acceptable hypotheses (18 to 26); in the latter case, only the best hypotheses constituted the feedback (6 to 7 ideas). The control group did not receive any feedback. The Formulating Hypotheses test represented the dependent variable, and its scoring was expanded to include a rated quality-score and a measure of the average number of words per response. Highly significant effects of the treatments on performance were obtained. Relative to the control group, the quantity model increased the number of responses and decreased the average number of words per response; the quality model increased the rated quality of the responses and the average number of

words per response but decreased the average number of responses. Of the two tests from the Kit of Reference Tests for Cognitive Factors (French et al. 1963) administered, Themes (ideational fluency) was significantly related to the number of responses and Advanced Vocabulary was significantly related to the rated quality of the responses. In their conclusion, Frederiksen and Evans expressed considerable doubt that the experimental treatments altered the participants' ability to formulate hypotheses. Rather, they maintained that the quantity and quality treatments simply changed participants' standards regarding a satisfactory performance.

Expansion of research on scientific thinking can be seen in the Frederiksen and Ward (1978) study, where measures extending beyond the Formulating Hypotheses test were developed. The general intent was to develop a set of measures that would have the potential to elicit creative scientific thinking while possessing psychometric acceptability. The authors sought to construct assessment devices in a middle ground between Guilford-type divergent-thinking tests (Guilford and Christensen 1956) and the global-creativity peer nominations of professional groups, typical of the work of MacKinnon (1962) and his collaborators. Leaning on the Flanagan (1949) study of critical incidents typical of scientists at work, Frederiksen and Ward attempted to develop instruments, called Tests of Scientific Thinking (TST), that would reflect problems that scientists often encounter in their work. The TST consisted of the Formulating Hypotheses test and three newly constructed tests: (a) Evaluating Proposals—test takers assume the role of an instructor and offer critical comments about proposals written by their students in a hypothetical science course; (b) Solving Methodological Problems—test takers offer solutions to a methodological problem encountered in planning a research study; and (c) Measuring Constructs—test takers suggest methods for eliciting relevant behavior for a specific psychological construct without resorting to ratings or self-reports. Scores tapped the quantity and quality of responses (statistical infrequency and ratings of especially high quality).

The TST was administered to students taking the *GRE*® Advanced Psychology Test. High levels of agreement prevailed among the four judges in scoring responses. However, the intercorrelations among the four tests varied considerably in magnitude, and Frederiksen and Ward (1978) concluded that there was "little evidence of generalized ability to produce ideas which are either numerous or good" (p. 11). It is to be expected, then, that factor analysis of the TST would yield multiple factors. A three-factor solution did in fact emerge, with Factor I reflecting the total number of responses and number of unusual responses, Factor II as a quality factor for Formulating Hypotheses and Measuring Constructs, and Factor III as a quality factor for Evaluating Proposals and Solving Methodological Problems. The Factor II tests were more divergent and imposed fewer constraints on the participants than did Factor III tests, which emphasized issues of design and analysis of experiments. The total number of responses and number of unusual responses cohering on Factor I parallels the findings with divergent-thinking tests where the number of unusual responses derives from the rate at which more obvious possibilities are exhausted. The factor analysis also makes clear that idea quality is unrelated to the number of proposed solutions.

Finally, Frederiksen and Ward (1978) inquired into the possible predictive validity of a composite of the four TSTs. A subgroup of the original sample, at the end of their first year in a graduate psychology-program, filled out a questionnaire with items inquiring into professional activities and accomplishments. Surprisingly, the scores for the number of responses from the TST composite yielded more significant relations with the questionnaire items than did the quality scores. Higher numbers of responses (mundane, unusual, and unusual high quality) were predictive of higher department quality, planning to work toward a Ph.D. rather than an M.A., generating more publications, engaging in collaborative research, and working with equipment. An inverse relation was found for enrollment in a program emphasizing the practice of psychology and for self-rated clinical ability. These outcomes strongly suggest that the TST may have value in forecasting the eventual productivity of a psychological scientist.

Two additional studies by ETS scientist Randy Bennett and his colleagues shed light on the validity of a computer-delivered Formulating Hypotheses test, which requires only general knowledge about the world, for graduate students from a variety of disciplines. Bennett and Rock (1995) used two four-item Formulating Hypotheses tests, one limiting the test takers' to seven-word responses and the other to 15-word responses. The tests were scored simply for the number of plausible, unduplicated hypotheses, based on the Frederiksen and Ward (1978) finding that the number of hypotheses is more highly related to criteria than their quality. A generalizability analysis showed high interjudge reliability. Generalizability coefficients for the mean ratings taken across judges and items were .93 for the seven-word version and .90 for the 15-word version. Three factors were identified in a confirmatory factor analysis of the two forms of the Formulating Hypotheses test and an ideational-fluency test (one item each from the Topics test of the Kit of Reference Tests for Cognitive Factors, French et al. 1963; the verbal form of the Torrance Tests of Creative Thinking, Torrance 1974; and two pattern-meaning tasks from the Wallach and Kogan 1965a, study). One factor was defined by the seven-word version, another by the 15-word version, and the third by the ideational-fluency test. The two formulating hypotheses factors correlated .90 with each other and .66 and .71 with the ideational-fluency factor. Bennett and Rock concluded that "the correlations between the formulating hypotheses factors …, though quite high, may not be sufficient to consider the item types equivalent" (p. 29).

Bennett and Rock (1995) also investigated the correlations of the two Formulating Hypotheses tests and the GRE General Test with two criterion measures: undergraduate grades and a questionnaire about extracurricular accomplishments in the college years (Stricker and Rock 2001), similar to the Baird (1979) and Skager et al. (1965) measures. The two tests had generally similar correlations with grades (r = .20 to .26 for the Formulating Hypotheses tests and .26 to .37 for the GRE General Test). The correlations were uniformly low between the tests and the six scales on the accomplishments questionnaire (Academic Achievement, Leadership, Practical Language [public speaking, journalism], Aesthetic Expression [creative writing, art, music, dramatics], Science, and Mechanical). Both Formulating Hypotheses tests correlated significantly with one of the scales: Aesthetic

Expression; and at least one of the GRE General Test sections correlated significantly with three scales: Aesthetic Expression, Academic Achievement, and Science.

The related issue of the Formulating Hypotheses test's incremental validity against these criteria was examined as well. The 15-word version of the test showed significant (but modest) incremental validity (vis-à-vis the GRE General Test) against grades (R^2 increased from .14 to .16). This version also demonstrated significant (but equally modest) incremental validity (vis-à-vis the GRE General Test and grades) against one of the six accomplishments scales: Aesthetic Expression (R^2 increased from .01 to .03). The seven-word version had no significant incremental validity against grades or accomplishments.

Enright et al. (1998), in a study to evaluate the potential of a Formulating Hypotheses test and experimental tests of reasoning for inclusion in the GRE General Test, replicated and extended the Bennett and Rock (1995) investigation of the Formulating Hypotheses tests. Enright et al. used the Bennett and Rock (1995) 15-word version of the test (renamed Generating Explanations), scored the same way. Four factors emerged in a confirmatory factor-analysis of the test with the GRE General Test's Verbal, Quantitative, and Analytical sections, and the three reasoning tests. The factors were Verbal, defined by the Verbal section, all the reasoning tests, and the logical-reasoning items from the Analytical section; Quantitative, defined only by the Quantitative section; Analytical, defined only by the analytical- reasoning items from the Analytical section; and Formulating Hypotheses, defined only by the Formulating Hypotheses test. The Formulating Hypotheses factor correlated .23 to .40 with the others.

Like Bennett and Rock (1995), Enright et al. (1998) examined the correlations of the Formulating Hypotheses test and the GRE General Test with undergraduate grades and accomplishments criteria. The Formulating Hypotheses test had lower correlations with grades ($r = .15$) than did the GRE General Test ($r = .22$ to .29). The two tests had consistently low correlations with the same accomplishments questionnaire (Stricker and Rock 2001) used by Bennett and Rock. The Formulating Hypotheses test correlated significantly with the Aesthetic Expression and Practical Language scales, and a single GRE General Test section correlated significantly with the Academic Achievement, Mechanical, and Science scales.

Enright et al. (1998) also looked into the incremental validity of the Formulating Hypotheses Test against these criteria. The test's incremental validity (vis-à-vis the GRE General Test) against grades was not significant for the total sample, but it was significant for the subsample of humanities and social-science majors (the increment was small, with R^2 increasing from .12 to .16). Enright et al. noted that the latter result is consistent with the test's demonstrated incremental validity for the total sample in the Bennett and Rock (1995) study, for over 60% of that sample were humanities and social-science majors. The test had no significant incremental validity (vis-à-vis the GRE General Test and grades) against an overall measure of accomplishments (pooling accomplishments across six areas), perhaps because of the latter's heterogeneity.

To sum up, the Bennett and Rock (1995) and Enright et al. (1998) investigations are remarkably consistent in demonstrating the distinctiveness of Formulating

Hypotheses tests from the GRE General Test and suggesting that the former can make a contribution in predicting important criteria.

In all of the TST research, a free-response format had been employed. The Formulating Hypotheses test lends itself to a machine-scorable version, and Ward et al. (1980) examined the equivalence of the two formats. In the machine-scorable version, nine possible hypotheses were provided, and the test taker was required to check those hypotheses that could account for the findings and to rank order them from best to worst. Comparable number and quality scores were derived from the two formats. The free-response/machine-scorable correlations ranged from .13 to .33 in a sample of undergraduate psychology majors, suggesting that the two versions were not alternate forms of the same test. When scores from the two versions were related to scores on the GRE Aptitude Test and the GRE Advanced Psychology Test, the correlations with the machine-scorable version were generally higher than those for the free-response version. Ward et al., in fact, suggested that the machine-scorable version offered little information beyond what is provided by the two GRE tests, whereas the free-response version did offer additional information. The obvious difference between the two versions is that the free-response requires test takers to produce solutions, whereas the machine-scorable merely calls for recognition of appropriate solutions. From the standpoint of ecological validity, it must be acknowledged that solutions to scientific problems rarely assume multiple-choice form. As Ward et al. point out, however, free-response tests are more difficult and time-consuming to develop and score, and yet are less reliable than multiple-choice tests of the same length.

14.1.3 Creative Products

Within a predictor-criterion framework, the previous two sections have focused on the former—individual differences in creative ability as reflected in performance on tests purportedly related to creativity on analogical or theoretical grounds. In some cases, various creativity criteria were available, making it possible to examine the concurrent or predictive validity of the creativity tests. Such research is informative about whether the creativity or scientific thinking label applied to the test is in fact warranted. In the present section, the focus is on the creative product itself. In some cases, investigators seek possible associations between the judged creativity of the product and the demographic or psychological characteristics of the individual who produced the product. In such instances, the predictor-to-criterion sequence is actually reversed.

Study of creative products can take two forms. The most direct form involves the evaluation of a concrete product for its creativity. A second and somewhat less direct form relies on test-takers' self-reports. The test taker is asked to describe his or her activities and accomplishments that may reflect different kinds of creative production, concrete or abstract, in such domains as science, literature, visual arts, and

music. It is the test-taker's verbal description of a product that is evaluated for creativity rather than the product itself.

14.1.3.1 Concrete Products

A good example of the concrete product approach to creativity is a study by ETS scientists Skager et al. (1966a). They raised the issue of the extent of agreement among a group of 28 judges (24 artists and 4 nonartists) in their aesthetic-quality ratings of drawings produced by the 191 students in the sophomore class at the Rhode Island School of Design. The students had the common assignment of drawing a nature scene from a vantage point overlooking the city of Providence. The question was whether the level of agreement among the judges in their quality ratings would be so high as to leave little interjudge variance remaining to be explained. This did not prove to be the case, as a varimax rotation of a principal-axis factor analysis of the intercorrelations of the 28 judges across 191 drawings suggested that at least four points of view about quality were discernible. Different artist judges were located on the first three factors, and the nonartists fell on the fourth factor. Factor I clearly pointed to a contrast between judges who preferred more unconventional, humorous, and spontaneous drawings versus judges who favored more organized, static, and deliberate drawings. Factors II and III were not readily distinguished by drawing styles, but the nonartists of Factor IV clearly expressed a preference for drawings of a more deliberate, less spontaneous style. Skager et al. next turned to the characteristics of the students producing the drawings and whether these characteristics might relate to the location of the drawing on one of the four quality points of view. Correlations were reported between the points of view and the students' scores on a battery of cognitive tests as well as on measures of academic performance, cultural background, and socioeconomic status. Most of these correlations were quite low, but several were sufficiently intriguing to warrant additional study, notably, majoring in fine arts, cultural background, and socioeconomic status.

Further analysis of the Skager et al. (1966a) data is described in Klein and Skager (1967). Drawings with the highest positive and negative factor loadings on the first two factors extracted in the Skager et al. study (80 drawings in all) were selected with the aim of further clarifying the spontaneous-deliberate contrast cited earlier. Ten lay judges were given detailed definitions of spontaneity and deliberateness in drawing and were asked to classify the drawings by placing each of them in a spontaneous or deliberate pile. A three-dimensional (judges × viewpoint × high vs. low quality) contingency table was constructed, and its chi-square was partitioned to yield main and interaction effects. A highly significant viewpoint × quality interaction was found. High-quality drawings for both Factor I and Factor II viewpoints were more likely to be classified as spontaneous relative to low-quality drawings. However, the effect was much stronger for the Factor I viewpoint, thereby accounting for the significant interaction. In sum, for Factor I judges, spontaneity versus deliberateness was a key dimension in evaluating the quality of the drawings; Factor

II judges, on the other hand, were evidently basing their evaluations on a dimension relatively independent of spontaneity-deliberateness. Of further interest is the extent to which lay judges, although differing from art experts on what constitutes a good drawing, nevertheless can, with minimal instruction, virtually replicate the aesthetic judgments of art experts holding a particular viewpoint (Factor I). These findings point to the potential efficacy of art appreciation courses in teaching and learning about aesthetic quality.

In a third and final approach to the topic of judged aesthetic quality of drawings, Skager et al. (1966b) subjected their set of 191 drawings to multidimensional scaling (MDS). For this purpose, 26 judges were selected from the faculty of nine schools of design across the United States. Because the scaling procedure required similarity ratings in paired comparisons of an entire stimulus set, practicality required that the size of the set be reduced. Accordingly, 46 of the 191 drawings were selected, reflecting a broad range in aesthetic quality as determined in prior research, and the 46 was divided into two equally sized subsets. Three dimensions emerged from separate MDS analyses of the two subsets. When factor scores for these dimensions were correlated with the test scores and other measures (the same battery used in the Skager et al. 1966a, study), the corresponding correlations in the two analyses correlated .64, suggesting that the three dimensions emerging from the two analyses were reasonably comparable.

What was the nature of the three dimensions? Skager et al. (1966b) chose to answer this question by comparing the drawings with the highest and lowest scale values on each dimension. There is a subjective and impressionistic quality to this type of analysis, but the outcomes were quite informative nevertheless. Dimension I offered a contrast between relative simplification versus complexity of treatment. The contrast was suggestive of the simplicity-complexity dimension described by Barron (1953). Among the contrasts distinguishing Dimension II were little versus extensive use of detail and objects nearly obscured versus clearly delineated. There was no obvious label for these contrasts. Dimension III contrasted neatness versus carelessness and controlled versus chaotic execution. The correlations between the three dimensions, on the one hand, and the test scores and other measures, on the other, were low or inconsistent between the two analyses. Skager et al. noted that some of the contrasts in the drawing dimensions carried a personality connotation (e.g., impulsiveness vs. conscientiousness). Because no personality tests were administered to the students who produced the drawings, this interesting speculation about the basis for aesthetic preference could not be verified.

Finally, we consider two studies by Ward and Cox (1974) that explored creativity in a community sample. Listeners to a New York City radio station were invited to submit humorous and original small green objects to a Little Green Things contest, with a reward of $300 for the best entry and 10 consolation prizes of $20 each. In the first study, a total of 283 submitted objects were rated for originality on a seven-point scale by four judges. The rated originality of each object represented the average of the judges' ratings, yielding an interjudge reliability of .80. Some 58% of the objects were found things, 34% were made things, and 8% were verbal additions to found things. Names and addresses of the contestants made it possible to determine

their gender and the census tract in which they resided. From the latter, estimates of family income and years of schooling could be derived. These demographic characteristics of the contestants could then be related to the originality ratings of their submissions. For all of the entries, these correlations were close to zero, but separating out made things yielded significant positive correlations of originality with estimates of family income and years of schooling. Unlike the case for verbal-symbolic forms of creativity assessment, where associations with socioeconomic status have generally not been found, a nonlaboratory context seemed to elevate the importance of this variable. Of course, this relationship occurred for made things—where some investment of effort was required. Why this should be so is unclear.

In a second study, using another set of objects, Ward and Cox (1974) attempted to uncover the dimensions possibly underlying the global originality rating of the objects. Judges were asked to rate the attractiveness, humor, complexity, infrequency, and effort involved in securing or making the object. A multiple *R* was computed indicating how much these five dimensions contributed to the object's originality rating. For found things, *R*s ranged from .25 to .73 (median = .53) with infrequency the strongest and humor the next strongest contributor; for made things, *R*s ranged from .47 to .71 (median = .64) with humor the strongest and amount of effort the next strongest contributor. It should be emphasized that the judges' evaluations were multidimensional, and for virtually every judge a combination of predictors accounted for more variance in the originality ratings than did any single predictor.

14.1.3.2 Reports of Products

A study by Skager et al. (1965) exemplifies the reporting of products approach. Using samples of college freshmen drawn from two institutions of higher learning (a technical institute and a state university), Skager et al. employed the Independent Activities Questionnaire, modeled after one devised by Holland (1961) and covering creative accomplishments outside of school during the secondary-school years. A sample item: "Have you ever won a prize or award for some type of original art work?" The number of these accomplishments served as a quantity score. Judges examined the participant's brief description of these activities with the goal of selecting the most significant achievement. These achievements were then given to a panel of judges to be rated on a 6-point scale to generate a quality score.

Quantity and quality scores were significantly correlated (*r*'s of .44 and .29). In a certain respect, this correlation is analogous to the significant correlations found between the fluency and uniqueness scores derived from ideational-fluency tests. The more divergent-thinking responses ventured or extracurricular activities undertaken by an individual, the more likely an original idea or high-quality accomplishment, respectively, will ensue. In neither sample did the quantity or quality scores relate to socioeconomic status, SAT Verbal, SAT Math, or high-school rank. The quantity score, however, did relate significantly in both samples with "an estimate from the student of the number of hours spent in discussing topics 'such as scientific

issues, world affairs, art, literature, or drama' with adults living in the home" (Skager et al. 1965, p. 34). By combining the samples from the two institutions, the quality score began to show significant relationships with SAT Verbal and SAT Math. This result simply reflected the enhanced variance in SAT scores and is of greater methodological than substantive interest.

ETS scientists Baird and Knapp (1981) carried out a similar study with the Inventory of Documented Accomplishments (Baird 1979), devised for graduate school. The inventory, concerning extracurricular accomplishments in the college years, had four scales measuring the number of accomplishments in these areas: literary-expressive, artistic, scientific-technical, social service and organizational activity. It was administered to incoming, first-year graduate students in English, biology, and psychology departments. At the end of their first year, the students completed a follow-up questionnaire about their professional activities and accomplishments in graduate school. The four scales correlated significantly with almost all of these activities and accomplishments, though only one correlation exceeded .30 (r = .50 for the Scientific-Technical scale with working with equipment). Because the sample combined students from different fields, potentially distorting these correlations, the corresponding correlations within fields were explored. Most of the correlations were higher than those for the combined sample.

14.1.4 Overview

Creativity research has evolved since the heyday of ETS's efforts in the 1960s and 1970s, at the dawn of psychology's interest in this phenomenon. The first journal devoted to creativity, the *Journal of Creative Behavior*, was published in 1967, followed by others, notably the *Creativity Research Journal*; *Psychology of Aesthetics, Creativity, and the Arts;* and *Imagination, Creativity, and Personality.* Several handbooks have also appeared, beginning with the Glover et al. (1989) *Handbook of Creativity* (others are Kaufman and Sternberg 2006, 2010b; Sternberg 1999; Thomas and Chan 2013). The volume of publications has burgeoned from approximately 400 articles before 1962 (Taylor and Barron 1963) to more than 10,000 between 1999 and 2010 (Kaufman and Sternberg 2010a). And the research has broadened enormously, "a virtual explosion of topics, perspectives, and methodologies…." (Hennessey and Amabile 2010, p. 571).

Nonetheless, divergent-thinking tests, evaluations of products, and inventories of accomplishments, the focus of much of the ETS work, continue to be mainstays in appraising individual differences. Divergent-thinking tests remain controversial (Plucker and Makel 2010), as noted earlier. The evaluation of products, considered to be the gold standard (Plucker and Makel 2010), has been essentially codified by the wide use of the Consensus Assessment Technique (Amabile 1982), which neatly skirts the knotty problem of defining creativity by relying on expert judges' own implicit conceptions of it. And there now seems to be a consensus that inventories of accomplishments, which have proliferated (see Hovecar and Bachelor 1989;

Plucker and Makel 2010), are the most practical and effective assessment method (Hovecar and Bachelor 1989; Plucker 1990; Wallach 1976).

Creativity is not currently an active area of research at ETS, but its earlier work continues to have an influence on the field. According to the Social Science Citation Index, Wallach and Kogan's 1965 monograph, *Modes of Thinking in Young Children*, has been cited 769 times through 2014, making it a citation classic.

14.2 Cognitive Styles

Defined as individual differences in ways of organizing and processing information (or as individual variation in modes of perceiving, remembering, or thinking), cognitive styles represented a dominant feature of the ETS research landscape beginning in the late 1950s and extending well into the 1990s. The key players were Samuel Messick, Kogan, and Herman Witkin, along with his longtime collaborators, Donald Goodenough and Philip Oltman, and the best known style investigated was field dependence-independence (e.g., Witkin and Goodenough 1981). The impetus came from Messick, who had spent a postdoctoral year at the Menninger Foundation (then a center for cognitive-style research) before assuming the leadership of the personality research group at ETS. During his postdoctoral year at Menninger, Messick joined a group of researchers working within an ego-psychoanalytic tradition who sought to derive a set of cognitive constructs that mediated between motivational drives and situational requirements. These constructs—six in all—were assigned the label of cognitive-control principles and were assessed with diverse tasks in the domains of perception (field dependence-independence), attention (scanning), memory (leveling-sharpening), conceptualizing (conceptual differentiation), susceptibility to distraction and interference (constricted-flexible control), and tolerance for incongruent or unrealistic experience (Gardner et al. 1959). Messick's initial contribution to this effort explored links between these cognitive-control principles and traditional intellectual abilities (Gardner et al. 1960). This study initiated the examination of the style-ability contrast—whereas abilities almost always reflect maximal performance, styles generally tap typical performance.

The psychoanalytic origin of the cognitive-control principles accounts for the emphasis on links to drives and defenses in early theorizing, but later research and theory shifted to the study of the cognitive-control principles (relabeled cognitive styles) in their own right. Messick played a major role in this effort, launching a project, supported by the National Institute of Mental Health, focused on conceptual and measurement issues posed by the assessment of these new constructs. The project supported a series of empirical contributions as well as theoretical essays and scholarly reviews of the accumulating literature on the topic. In this effort, Messick was joined by Kogan, who collaborated on several of the empirical studies—conceptual differentiation (Messick and Kogan 1963), breadth of categorization and quantitative aptitude (Messick and Kogan 1965), and a MDS approach to cognitive

complexity-simplicity (Messick and Kogan 1966). Other empirical work included studies of the influence of field dependence on memory by Messick and Damarin (1964) and Messick and Fritzky (1963). Scholarly reviews were published (Kagan and Kogan 1970; Kogan 1971) that enhanced the visibility of the construct of cognitive style within the broader psychological and educational community. Messick (1970) provided definitions for a total of nine cognitive styles, but this number expanded to 19 six years later (Messick 1976). It is evident that Messick's interest in cognitive styles at that latter point in time had moved well beyond his original psychoanalytic perspective to encompass cognitive styles generated by a diversity of conceptual traditions.

The reputation of ETS as a center for cognitive-style research was further reinforced by the 1973 arrival of Witkin, with Goodenough and Oltman. This team focused on field dependence-independence and its many ramifications. A field-independent person is described as able to separate a part from a whole in which it is embedded—the simple figure from the complex design in the Embedded Figures Test (EFT) and the rod from the tilted frame in the Rod and Frame Test (RFT). A field-dependent person is presumed to find it difficult to disembed part of a field from its embedding context. The Witkin team was exceptionally productive, generating empirical studies (e.g., Witkin et al. 1974, 1977a; Zoccolotti and Oltman 1978) and reviews (e.g., Goodenough 1976; Witkin and Berry 1975; Witkin and Goodenough 1977; Witkin et al. 1977b, 1979) that stamped Witkin as one of the foremost personality researchers of his era (Kogan 1980). His death in 1979 severely slowed the momentum of the field dependence-independence enterprise to the point where its future long-term viability was called into question. Nevertheless, further conceptual and methodological refinement of this construct continued in articles published by Messick (1984, 1987, 1994, 1996) and in empirical work and further conceptualizing by Goodenough and his colleagues (e.g., Goodenough 1981, 1986; Goodenough et al. 1987, 1991).

Kogan, who had by then departed for the New School for Social Research, continued to build upon his ETS experience and devoted several publications to field dependence-independence and other cognitive styles (Kogan 1976, 1983, 1994; Kogan and Saarni 1990). A conference bringing together the principal field dependence-independence theorists and researchers (domestic and foreign) was held at Clark University in 1989 and subsequently appeared as an edited book (Wapner and Demick 1991). Kogan and Block (1991) contributed a chapter to that volume on the personality and socialization aspects of field dependence-independence. That chapter served to resolve conceptual incongruities that arose when the Witkin team altered their original value-laden theory (Witkin et al. 1962) in a direction favoring a value-neutral formulation (Witkin and Goodenough 1981). The latter endowed field-dependent and field-independent individuals with distinctive sets of skills—analytic restructuring versus interpersonal, respectively. The extensive longitudinal research reported by Kogan and Block proved more consistent with the earlier formulation (Witkin et al. 1962) than with the more recent one (Witkin and Goodenough 1981).

Educational implications of cognitive styles were of particular interest at ETS, and ETS researchers made contributions along those lines. Working with the nine cognitive styles delineated by Messick (1970), a book chapter by Kogan (1971) pointed to much variation at that point in time in the degree to which empirical investigations based on those styles could be said to offer implications for education. Indeed, for some of the styles, no effort had been made to establish educational linkages, not surprising given that the origins of cognitive styles can be traced to laboratory-based research on personality and cognition. It took some years before the possibility of educational applications received any attention. By the time of a subsequent review by Kogan (1983), this dearth had been corrected, thanks in large part to the work of Witkin and his colleagues, and subsequently to Messick's (1984, 1987) persistent arguments for the importance of cognitive styles in accounting for educational processes and outcomes. Witkin and his colleagues considered the educational implications of field dependence-independence in a general survey of the field (Witkin et al. 1977b) and in an empirical study of the association between field dependence-independence and college students' fields of concentration (Witkin et al. 1977a). In the latter study, three broad categories of student majors were formed: (a) science; (b) social science, humanities, and arts; and (c) education. Field independence, assessed by EFT scores, was highest for science majors and lowest for education majors. Furthermore, students switching out of science were more field dependent than those who remained, whereas students switching out of education were more field independent than those who remained. An attempt to relate field dependence-independence to performance (i.e., grades) within each field yielded only marginal results. The findings clearly supported the relevance of field dependence-independence as an important educational issue.

Another topic of educational relevance is the matching hypothesis initially framed by Cronbach and Snow (1977) as a problem in aptitude-treatment interaction. The basic proposition of this interaction is that differences among learners (whether in aptitude, style, strategy, or noncognitive attributes) may imply that training agents or instructional methods can be varied to capitalize upon learners' strengths or to compensate for their weaknesses. A portion of this research inquired into cognitive styles as individual differences in the aptitude-treatment interaction framework, and the Witkin et al. (1977b) review showed inconsistent effects for field dependence-independence across several studies. There was some indication that style-matched teachers and students liked one another more than did mismatched pairs, but there was little evidence suggesting that matching led to improved learning outcomes. A more recent review by Davis (1991) of field dependence-independence studies of this kind again suggests a mixed picture of successes and failures. Messick (1994, 1996) has attributed many of these failures to the haphazard manner in which field dependence-independence has been assessed. Typically, the isolated use of the EFT to assess this cognitive style implies that only the cognitive restructuring or set-breaking component is represented in the field dependence-independence index to the exclusion of the component represented by the RFT, which Witkin and Goodenough (1981) described as visual versus vestibular sensitivity to perception of the upright. They, in fact, raised the possibility that the EFT

and RFT may be tapping distinctive, although related, psychological processes. Multivariate studies of a diversity of spatial tasks have found that EFT and RFT load on separate factors (Linn and Kyllonen 1981; Vernon 1972). A discussion of the implications of these findings can be found in Kogan (1983).

14.2.1 Conclusion

There can be no doubt that the field dependence-independence construct has faded from view, but this in no way implies that the broader domain of cognitive styles has correspondingly declined. More recently, Kozhevnikov (2007) offered a review of the cognitive-style literature that envisions the future development of theoretical models incorporating neuroscience and research in other psychological fields. Thus, the style label has been attached to research on decision-making styles (e.g., Kirton 1989), learning styles (e.g., Kolb 1976), and thinking styles (e.g., Sternberg 1997). Possibly, the dominant approach at present is that of intellectual styles (Zhang and Sternberg 2006). Zhang and Sternberg view intellectual styles as a very broad concept that more or less incorporates all prior concepts characterizing stylistic variation among individuals. This view will undoubtedly be reinforced by the recent appearance of the *Handbook of Intellectual Styles* (Zhang et al. 2011). It is dubious, however, that so broad and diverse a field as stylistic variation among individuals would be prepared to accept such an overarching concept at the present stage in its history.

14.3 Risk Taking

Research on risk-taking behavior, conducted by Kogan and Wallach, was a major activity at ETS in the 1960s. Despite the importance and general interest in this topic, no review of research in the field had been published prior to their essay (Kogan and Wallach 1967c). In that review, they surveyed research on situational, personal, and group influences on risk-taking behavior. Also discussed were the assets and liabilities of mathematical models (e.g., Edwards 1961; Pruitt 1962) developed to account for economic decision making and gambling behavior. Simon (1957) rejected this rational economic view of the individual as maximizer in favor of the individual as satisfier—accepting a course of action as good enough. This latter perspective on decision making is more friendly to the possibility of systematic individual-variation in what constitutes good enough, and thus opened the door to a construct of risk taking.

14.3.1 Individuals

In the matter of situational influences, the distinction between chance tasks and skill tasks would seem critical, but the contrast breaks down when taking account of the degree of control individuals believe they can exert over decision outcomes. In the Kogan and Wallach (1964) research, a direct comparison between risk preferences under chance and skill conditions was undertaken. Participants gambled for money on games of chance (e.g., dice) and skill (e.g., shuffleboard), choosing the bets and resultant monetary payoffs (the games scored for maximizing gains, minimizing losses, and deviations from a 50-50 bet). There was no indication in the data of greater risk taking under skill conditions. Rather, there was a strategic preference for moderate risk taking (minimizing deviation from a 50-50 bet). By contrast, the chance condition yielded greater variance as some participants leaned toward risky strategies or alternatively toward cautious strategies.

Variation in risk-taking behavior can also be observed in an information-seeking context. The paradigm is one in which there is a desirable goal (e.g., a monetary prize for solving a problem) with informational cues helpful to problem solution offered at a price. To avail oneself of all the cues provided would reduce the prize to a negligible value. Hence, the risk element enters as the person attempts to optimize the amount of information requested. Venturing a solution early in the informational sequence increases the probability of an incorrect solution that would forfeit the prize. Such a strategy is indicative of a disposition toward risk taking. Irwin and Smith (1957) employed this information-seeking paradigm and observed that the number of cues requested was directly related to the value of the prize and inversely related to the monetary cost per cue. Kogan and Wallach (1964) employed information-seeking tasks in their risk-taking project.

The risk-taking measures described thus far were laboratory based and cast decisions in a gambling-type format with monetary incentives (while avoiding use of participants' own money). Most real-life decision making does not conform to the gambling paradigm, and accordingly, Kogan and Wallach (1964) constructed a series of choice dilemmas drawn from conceivable events in a variety of life domains. An abbreviated version of a scenario illustrates the idea: "Mr. A, an electrical engineer, had the choice of sticking with his present job at a modest, though adequate salary, or of moving on to another job offering more money but no long-term security." These scenarios (12 in all) constituted the Choice Dilemmas Questionnaire (CDQ). In each of these scenarios, the participant is asked to imagine advising the protagonist, who is faced with the choice between a highly desirable alternative with severe negative consequences for failure and a less desirable alternative where consequences for failure are considerably less severe. On a probability scale extending from 9 in 10 to 1 in 10, the participant is asked to select the minimum odds of success the protagonist should demand before opting for the highly desirable alternative. Descending the probability scale (toward 1 in 10) implies increasing preference for risk. (A 10 in 10 option is also provided for participants demanding complete certainty that the desirable alternative will be successful.) The

CDQ has also been claimed to measure the deterrence value of potential failure in the pursuit of desirable goals (Wallach and Kogan 1961). Its reliability has ranged from the mid-.50s to the mid-.80s.

Diverse tasks have been employed in the assessment of risk-taking dispositions. The basic question posed by Kogan and Wallach (1964) was whether participants demonstrate any consistency in their risk-taking tendencies across these tasks. The evidence derived from samples of undergraduate men and women pointed to quite limited generality, calling into question the possibility of risk-inclined versus prudent, cautious personalities. A comparable lack of cross-situational consistency had been observed earlier by Slovic (1962). Unlike Slovic, however, Kogan and Wallach chose to explore the role of potential moderators selected for their conceptual relevance to the risk-taking domain. The first moderator considered was test anxiety. Atkinson (1957) conceptualized test anxiety as fear of failure and offered a model in which fear-of-failure individuals would make exceedingly cautious or risky choices in a level-of-aspiration problem-solving paradigm. Cautious choices are obviously more likely to ensure success, and exceptionally risky choices offer a convenient rationalization for failure. Hence, test-anxious participants were expected to be sensitized to the success and failure potentialities of the risk-taking measures with the likely consequence of enhanced consistency in their choices. The second moderator under examination was defensiveness—also labeled need for approval, by Crowne and Marlowe (1964). Many of the tasks employed in the Kogan and Wallach research required a one-on-one interaction with an experimenter. Participants high in defensiveness were considered likely to engage in impression management—a desire to portray oneself consistently as a bold decision-maker willing to take risks, or as a cautious, prudent decision-maker seeking to avoid failure. Accordingly, enhanced cross-task consistency was anticipated for the highly defensive participants.

Both moderators proved effective in demonstrating the heightened intertask consistency of the high test-anxious and high-defensive participants relative to the participants low on both moderators. The latter subgroup's risk-taking preferences appeared to vary across tasks contingent on their stimulus properties, whereas the former, motivationally disturbed subgroups appeared to be governed by their inner motivational dispositions in tasks with a salient risk component. It should be emphasized that Kogan and Wallach (1964) were not the first investigators to discover the value of moderator analyses in the personality domain. Saunders (1956) had earlier reported enhanced predictability through the use of personality moderators. More recently, Paunonen and Jackson (1985) offered a multiple-regression model for moderator analyses as a path toward a more idiographic approach in personality research.

Of further interest in the Kogan and Wallach (1964) monograph is the evidence indicating an association between risk-taking indices and performance on the SAT Verbal section for undergraduate men. The relationship was moderated by test anxiety such that high test-anxious participants manifested an inverse association and low test-anxious participants a direct association between risk-taking level and SAT performance. In short, a disposition toward risk taking facilitated the low-anxious

person (presumably enabling educated guessing) and hindered the anxiety-laden individual (presumably due to interference with cognitive processing). Hence, the penalty-for-guessing instructions for the SAT (retained by the College Board until recently) seemed to help some participants while hurting others.

Beyond the consistency of risk-taking dispositions in the motivationally-disturbed participants, Kogan and Wallach (1964) introduced the possibility of irrationality in the choices of those subgroups. After implementing their choices, participants were informed of their monetary winnings and offered the opportunity to make a final bet with those winnings on a single dice toss that could enhance those winnings up to six-fold if successful but with the risk of total loss if unsuccessful. The low anxious/low defensive participants exhibited the *protecting-one's-nest-egg* phenomenon in the sense of refusing to make a final bet or accepting less risk on the bet in proportion to the magnitude of their winnings. In the motivationally disturbed subgroups, on the other hand, the magnitude of winnings bore no relation to the risk level of the final bet. In other words, these subgroups maintained their consistently risky or cautious stance, essentially ignoring how much they had previously won. Further evidence for irrationality in the motivationally disturbed subgroups concerned post-decisional regret. Despite a frequent lack of success when playing their bets, participants in those subgroups expressed minimal regret about their original decisions unlike the low anxious/low defensive participants who wished they could alter original choices that failed to yield successful outcomes. In the sense that some participants ignored relevant situational properties whereas others took account of them, the issue of rationality-irrationality became germane.

The directions taken by risk-taking research subsequent to the Kogan and Wallach (1964, 1967c) contributions were summarized in the chapters of a book edited by Yates (1992). At that time, the issue of individual and situational influences on risk-taking preferences and behavior remained a focus of debate (Bromiley and Curley 1992). That risk taking continues as a hot topic is demonstrated in the research program undertaken by Figner and Weber (2011). They have introduced contrasts in the risk-taking domain that had received little attention earlier. For example, they distinguish between affective and deliberative risk-taking (also described as hot vs. cold risk-taking). Thus, a recreational context would be more likely to reflect the former, and a financial investment context the latter.

14.3.2 Small Groups

14.3.2.1 Intragroup Effects

It is often the case that major decisions are made, not by individuals acting alone, but by small groups of interacting individuals in an organizational context. Committees and panels are often formed to deal with problems arising in governmental, medical, and educational settings. Some of the decisions made by such groups entail risk assessments. The question then arises as to the nature of the

relationship between the risk level of the individuals constituting the group and the risk level of the consensus they manage to achieve. Most of the research directed to this issue has employed groups of previously unacquainted individuals assembled solely for the purpose of the experiments. Hence, generalizability to longer-term groups of acquainted individuals remains an open question. Nevertheless, it would be surprising if the processes observed in minimally acquainted groups had no relevance for acquainted individuals in groups of some duration.

There are three possibilities when comparing individual risk preferences with a consensus reached through group discussion. The most obvious possibility is that the consensus approximates the average of the prior individual decisions. Such an outcome obviously minimizes the concessions required of the individual group members (in shifting to the mean), and hence would seem to be an outcome for which the members would derive the greatest satisfaction. A second possible outcome is a shift toward caution. There is evidence that groups encourage greater carefulness and deliberation in their judgments, members not wishing to appear foolhardy in venturing an extreme opinion. The third possibility is a shift toward greater risk taking. There is mixed evidence about this shift from brainstorming in organizational problem-solving groups (Thibaut and Kelley 1959), and the excesses observed in crowds have been described by Le Bon (1895/1960). Both of these situations would seem to have limited relevance for decision-making in small discussion groups.

This third possibility did emerge in an initial study of decision-making in small discussion groups (Wallach et al. 1962). College students made individual decisions on the CDQ items and were then constituted as small groups to reach a consensus and make individual post-discussion decisions. The significant risky shifts were observed in the all-male and all-female groups, and for both the consensus and post-consensus-individual decisions. Interpretation of the findings stressed a mechanism of diffusion of responsibility whereby a group member could endorse a riskier decision because responsibility for failure would be shared by all of the members of the group.

It could be argued, of course, that decision making on the CDQ is hypothetical—no concrete payoffs or potential losses are involved—and that feature could account for the consistent shift in the risky direction. A second study (Wallach et al. 1964) was designed to counter that argument. SAT items of varying difficulty levels (10% to 90% failure rate as indicated by item statistics) were selected from old tests, and monetary payoffs were attached proportional to difficulty level to generate a set of choices equal in expected value. College students individually made their choices about the difficulty level of the items they would be given and then were formed into small groups with the understanding they would be given the opportunity to earn the payoff if the item was answered correctly. A risky shift was observed (selecting more difficult, higher payoff items) irrespective of whether responsibility for answering the selected item was assigned to a particular group member or to the group as a whole. The monetary prize in each case for a successful solution was made available to each group member. Again, in a decision context quite different

from the CDQ, group discussion to consensus yielded risky shifts that lent themselves to explanation in terms of diffusion of responsibility.

A partial replication of the foregoing study was carried out by Kogan and Carlson (1969). In addition to sampling college students, a sample of fourth graders and fifth graders was employed. Further, a condition of overt intragroup competition was added in which group members bid against one another to attempt more difficult items. Consistent with the Wallach et al. (1964) findings, risky shifts with group discussion to consensus were observed in the sample of college students. The competition condition did not yield risky shifts, and the outcomes for the elementary school sample were weaker and more ambiguous than those obtained for college students.

While the preceding two studies provided monetary payoffs contingent on problem solution, participants did not experience the prospect of losing their own money. To enhance the risk of genuinely aversive consequences, Bem et al. (1965) designed an experiment in which participants made choices that might lead to actual physical pain coupled with monetary loss. (In actuality, participants never endured these aversive effects, but they were unaware of this fact during the course of the experiment.) Participants were offered an opportunity to be in experiments that differed in the risks of aversive side effects from various forms of stimulation (e.g., olfactory, taste, movement). Monetary payoffs increased with the percentage of the population (10% to 90%) alleged to experience the aversive side effects. Again, discussion to consensus and private decisions following consensus demonstrated the risky-shift effect and hence provided additional evidence for a mechanism of responsibility diffusion.

With the indication that the risky-shift effect generalizes beyond the hypothetical decisions of the CDQ to such contexts as monetary gain and risk of painful side-effects, investigators returned to the CDQ to explore alternative interpretations for the effect with the knowledge that it is not unique to the CDQ. Thus, Wallach and Kogan (1965b) experimentally split apart the discussion and consensus components of the risky-shift effect. Discussion alone without the requirement of achieving a consensus generated risky shifts whose magnitude did not differ significantly from discussion with consensus. By contrast, the condition of consensus without discussion (a balloting procedure where group members were made aware of each other's decisions by the experimenter and cast as many ballots as necessary to achieve a consensus), yielded an averaging effect. It is thus apparent that actual verbal-interaction is essential for the risky shift to occur. The outcomes run contrary to Brown's (1965) interpretation that attributes the risky shift to the positive value of risk in our culture and the opportunity to learn in the discussion that other group members are willing to take greater risks than oneself. Hence, these members shift in a direction that yields the risky-shift effect. Yet, in the consensus-without-discussion condition in which group members became familiar with others' preferred risk levels, the outcome was an averaging rather than a risky-shift effect. Verbal interaction, on the other hand, not only allows information about others' preferences, but it also generates the cognitive and affective processes presumed necessary for responsibility diffusion to occur.

It could be contended that the balloting procedure omits the exchange of information that accompanies discussion, and it is the latter alone that might be sufficient to generate the risky-shift effect. A test of this hypothesis was carried out by Kogan and Wallach (1967d) who compared interacting and listening groups. Listeners were exposed to information about participants' risk preferences and to the pro and con arguments raised in the discussion as well. Both the interacting and listener groups manifested the risky-shift effect, but its magnitude was significantly smaller in the listening groups. Hence, the information-exchange hypothesis was not sufficient to account for the full strength of the effect. Even when group members were physically separated (visual cues removed) and communicated over an intercom system, the risky shift retained its full strength (Kogan and Wallach 1967a). Conceivably, the distinctiveness of individual voices and expressive styles allowed for the affective reactions presumed to underlie the mechanism of responsibility diffusion.

To what extent are group members aware that their consensus and individual post-consensus decisions are shifting toward greater risk-taking relative to their prior individual-decisions? Wallach et al. (1965) observed that group members' judgments were in the direction of shifts toward risk, but their estimates of the shifts significantly underestimated the actual shifts.

In a subsequent study, Wallach et al. (1968) inquired whether risk takers were more persuasive than their more cautious peers in group discussion. With risk-neutral material used for discussion, persuasiveness ratings were uncorrelated with risk-taking level for male participants and only weakly correlated for female participants. Overall, the results suggested that the risky shift could not be attributed to the greater persuasiveness of high risk takers. A different process seemed to be at work.

As indicated earlier, the paradigm employed in all of the previously cited research consisted of unacquainted individuals randomly assembled into small groups. Breaking with this paradigm, Kogan and Wallach (1967b) assembled homogeneous groups on the basis of participants' scores on test anxiety and defensiveness. Median splits generated four types of groups—high and/or low on the two dimensions. Both dimensions generated significant effects—test anxiety in the direction of a stronger risky shift and defensiveness in the direction of a weaker risky shift. These outcomes were consistent with a responsibility-diffusion interpretation. Test-anxious participants should be especially willing to diffuse responsibility so as to relieve the burden of possible failure. Defensive participants, by contrast, might be so guarded in relation to each other that the affective matrix essential for responsibility diffusion was hindered in its development.

In a related study, field dependence-independence served as the dimension for constructing homogeneous groups (Wallach et al. 1967). The magnitude of the risky shift was not significantly different between field-dependent and field-independent groups. There was a decision-time difference, with field-dependent groups arriving at a consensus significantly more quickly. The more time that was taken by field-dependent groups, the stronger the risky shift, whereas, the more time that was taken by field-independent groups, the weaker the risky shift. More time for field-dependent groups permitted affective bonds to develop, consistent with a process of

responsibility diffusion. More time for field-independent groups, by contrast, entailed resistance to other group members' risk preferences and extended cognitively based analysis, a process likely to mitigate responsibility diffusion.

A slight change in the wording of instructions on the CDQ transforms it from a measure of risk taking into a measure of pessimism-optimism. On a probability scale ranging from 0 in 10 to 10 in 10, the test taker is asked to estimate the odds that the risky alternative would lead to a successful outcome if chosen. Descending the probability scale (toward 1 in 10) implies increasing pessimism. Contrary to the expectation that a risky shift would lead to a surge of optimism, the outcome was a significant shift toward pessimism (Lamm et al. 1970). The discussion generated a consensus probability more pessimistic than the prediscussion average of the participating group members. When estimating success/failure probabilities, the discussion focused on things that might go wrong and the best alternative for avoiding error. Hence, the pessimism outcomes can be viewed as a possible constraint on extremity in risky decision-making.

14.3.2.2 Intergroup Effects

With financial support from the Advanced Research Projects Agency of the US Defense Department, Kogan and his collaborators undertook a series of studies in France, Germany, and the United States that departed from the standard intragroup-paradigm by adding an intergroup component. Participants in small decision-making groups were informed that one or more of them would serve as delegates meeting with delegates from other groups with the intent of presenting and defending the decisions made in their parent groups. Such a design has real-world parallels in the form of local committees arriving at decisions, where a representative is expected to defend the decisions before a broader-based body of representatives from other localities.

In an initial exploratory study with a French university sample (Kogan and Doise 1969), 10 of the 12 CDQ items with slight modifications proved to be appropriate in the French cultural context and were accordingly translated into French. Discussion to consensus on the first five CDQ items was followed by an anticipated delegate condition for the latter five CDQ items. Three delegate conditions were employed in which the group members were told (a) the delegate would be selected by chance, (b) the delegate would be selected by the group, and (c) all group members would serve as delegates. The significant shift toward risk was observed in the initial five CDQ items, and the magnitude of the risky shift remained essentially at the same level for all three of the anticipated delegate conditions. It is evident, then, that the expectation of possibly serving as a delegate in the future does not influence the processes responsible for the risky-shift effect.

In subsequent studies, delegates were given the opportunity to negotiate with each other. In the Hermann and Kogan (1968) investigation with American undergraduate men, dyads pairing an upperclassman (seniors, juniors) with an underclassman (sophomores, freshmen) engaged in discussion to consensus on the CDQ

items. The upperclassmen were designated as leaders, and the underclassmen as delegates. The risky shift prevailed at the dyadic level. Intergroup negotiation then followed among leaders and among delegates. The former manifested the risky shift, whereas the latter did not. This outcome is consistent with a responsibility-diffusion interpretation. Requiring delegates to report back to leaders would likely interfere with the affective processes presumed to underlie diffusion of responsibility. Leaders, by contrast, have less concern about reporting back to delegates. One cannot rule out loss-of-face motivation, however, and the magnitude of the risky shift in the leader groups was in fact weaker than that observed in the typical intragroup setting.

A follow-up to this study was carried out by Lamm and Kogan (1970) with a sample of German undergraduate men. As in the case of the French study (Kogan and Doise 1969), 10 of the 12 CDQ items (with slight modification) were considered appropriate in the German cultural context and were accordingly translated into German. Unlike the Hermann and Kogan (1968) study in which status was ascribed, the present investigation was based on achieved status. Participants in three-person groups designated a representative and an alternate, leaving a third individual designated as a nonrepresentative. Contrary to the Hermann and Kogan (1968) findings where leaders manifested the risky shift, the representative groups (presumed analogous to the leaders) failed to demonstrate the risky shift. On the other hand, the alternate and nonrepresentative groups did generate significant risky shifts. The argument here is that achieved, as opposed to ascribed, status enhanced loss-of-face motivation, making difficult the concessions and departures from prior intragroup decisions that are essential for risky shifts to occur. Having been assigned secondary status by the group, the alternates and nonrepresentatives were less susceptible to loss-of-face pressures and could negotiate more flexibly with their status peers.

In a third and final study of the delegation process, Kogan et al. (1972) assigned leader and subordinate roles on a random basis to German undergraduate men. The resultant dyads discussed the CDQ items to consensus (revealing the anticipated risky shift) and were assigned negotiating and observer roles in groups comprised exclusively of negotiators or observers. All four group types—leader-negotiators, subordinate-observers, subordinate-negotiators, and leader-observers—demonstrated the risky shift. However, the subordinate observers relative to their negotiating leaders preferred larger shifts toward risk. Evidently, loss-of-face motivation in the leaders in the presence of their subordinates served as a brake on willingness to shift from their initial-dyadic decisions. The nature of the arguments, however, convinced the observing subordinates of the merits of enhanced risk taking.

Two studies were conducted to examine preferred risk-levels when decisions are made for others. The first (Zaleska and Kogan 1971) utilized a sample of French undergraduate women selecting preferred probability and monetary stake levels in a series of equal-expected-value chance bets to be played for the monetary amounts involved. In addition to a control condition (self-choices on two occasions), three experimental conditions were employed: (a) individual choices for self and another, (b) individual and group choices for self, and (c) individual and group choices for

others. The first condition generated cautious shifts, the second yielded risky shifts, and the third produced weakened risky-shifts. Evidently, making individual choices for another person enhances caution, but when such choices are made in a group, a significant risky shift ensues, though weaker than obtained in the standard intra-group condition.

The findings bear directly on competing interpretations of the risky-shift effect. The popular alternative to the responsibility-diffusion interpretation is the risk-as-value interpretation initially advanced by Brown (1965) and already described. As noted in the Zaleska and Kogan (1971) study, caution is a value for individuals making choices for others, yet when deciding for others as a group, the decisions shifted toward risk. Such an outcome is consistent with a responsibility-diffusion interpretation, but the lesser strength of the effect suggests that the value aspect exerts some influence. Hence, the two conflicting interpretations may not necessarily assume an either-or form. Rather, the psychological processes represented in the two interpretations may operate simultaneously, or one or the other may be more influential depending on the decision-making context.

Choices in the Zaleska and Kogan (1971) study were distinguished by reciprocity—individuals and groups choosing for unacquainted specific others were aware that those others would at the same time be choosing for them. A subsequent study by Teger and Kogan (1975), using the Zaleska and Kogan chance bets task, explored this reciprocity feature by contrasting gambling choices made under reciprocal versus nonreciprocal conditions in a sample of American undergraduate women. A significantly higher level of caution was observed in the reciprocal condition relative to the nonreciprocal condition. This difference was most pronounced for high-risk bets that could entail possible substantial loss for the reciprocating other. Hence, the enhanced caution with reciprocity was most likely intended to ensure at least a modest payoff for another who might benefit the self. Caution in such circumstances serves the function of guilt avoidance.

We might ask whether the research on group risk-taking represented a passing fad. The answer is no. The group risk-taking research led directly to the study of polarization in small groups—the tendency for group discussion on almost any attitudinal topic to move participants to adopt more extreme positions at either pole (e.g., Myers and Lamm 1976). This polarization work eventually led to the examination of the role of majorities and minorities in influencing group decisions (e.g., Moscovici and Doise 1994). In short, the dormant group-dynamics tradition in social psychology was invigorated. Reviewing the group risk-taking research 20 years after its surge, Davis et al. (1992) noted that the "decline of interest in investigating the parameters of group risk taking was unfortunate" (p. 170). They go on to note the many settings in which group decision-making takes place (e.g., parole boards, juries, tenure committees) and where the "conventional wisdom persists that group decisions are generally moderate rather than extreme, despite such contrary evidence as we have discussed above" (p. 170).

14.4 Kinesthetic Aftereffect

A phenomenon originally demonstrated by Köhler and Dinnerstein in 1947, the kinesthetic aftereffect captured the attention of psychologists for almost a half-century. Early interest in this phenomenon can be traced to experimental psychologists studying perception who sought to establish its parameters. In due course, individual differences in the kinesthetic aftereffect attracted personality psychologists who viewed it as a manifestation of the augmenter-reducer dimension, which distinguishes between people who reduce the subjective intensity to external stimuli and those who magnify it (Petrie 1967).

Consider the nature of the kinesthetic-aftereffect task. A blindfolded participant is handed (right hand) a wooden test block 2 inches in width and 6 inches in length. The participant then is requested to match the width of the test block on an adjustable wedge (30 inches long) located to the participant's left hand. This process constitutes the preinduction measurement. Next, the participant is handed an induction block 1/2 inch narrower or wider than the test block and asked to give it a back-and-forth rubbing. Then the participant returns to the test block, and the initial measurement is repeated. The preinduction versus postinduction difference in the width estimate constitutes the score.

Kinesthetic-aftereffect research at ETS was conducted by A. Harvey Baker and his colleagues. One question that they examined was the effect of experimental variations on the basic paradigm just described. Weintraub et al. (1973) had explored the contrast between a wider and narrower induction block (relative to the test block) on the magnitude and direction of the kinesthetic aftereffect. They also included a control condition eliminating the induction block that essentially reduced the score to zero. The kinesthetic aftereffect proved stronger with the wider induction block, probably the reason that subsequent research predominantly employed a wider induction block, too.

Taking issue with the absence of an appropriate control for the induction block in the kinesthetic-affereffect paradigm, Baker et al. (1986) included a condition in which the test and induction blocks were equal in size. Such a control permitted them to determine whether the unequal size of the two blocks was critical for the kinesthetic aftereffect. Both the induction > test and induction < test conditions generated a significant kinesthetic aftereffect. The induction = test condition also yielded a significant kinesthetic aftereffect, but it was not significantly different from the induction > test condition. On this basis, Baker et al. concluded that two processes rather than one are necessary to account for the kinesthetic-aftereffect phenomenon—induction (rubbing the induction block) and the size contrast. It should be noted that these findings were published as research on this phenomenon had begun to wane, and hence their influence was negligible.

Two additional questions investigated by Baker and his coworkers in other research were the kinesthetic aftereffect's reliability and personality correlates. A stumbling block in research on this phenomenon was the evidence of low test-retest reliability across a series of trials. Until the reliability issue could be resolved, the

prospect for the kinesthetic aftereffect as an individual-differences construct remained dubious. Baker et al. (1976, 1978) maintained that test-retest reliability is inappropriate for the kinesthetic aftereffect. They noted that the kinesthetic aftereffect is subject to practice effects, such that the first preinduction-postinduction pairing changes the person to the extent that the second such pairing is no longer measuring the same phenomenon. In support of this argument, Baker et al. (1976, 1978) reviewed research based on a single-session versus a multiple-session kinesthetic aftereffect and reported that it was only the former that yielded significant validity coefficients with theoretically-relevant variables such as activity level and sensation seeking.

In another article on this topic, Mishara and Baker (1978) argued that internal-consistency reliability is most relevant for the kinesthetic aftereffect. Of the 10 samples studied, the first five employed the Petrie (1967) procedure in which a 45-minute rest period preceded kinesthetic-aftereffect measurements. Participants were not allowed to touch anything with their thumbs and forefingers during this period, and the experimenter used the time to administer questionnaires orally. The remaining five samples were tested with the Weintraub et al. (1973) procedure that did not employ the 45-minute rest period. For the samples using the Petrie procedure, the split-half reliabilities ranged from .92 to .97. For the samples tested with the Weintraub et al. procedure, the reliabilities ranged from .60 to .77. Mishara and Baker noted that the Weintraub et al. procedure employed fewer trials, but application of the Spearman-Brown correction to equate the number of trials in the Weintraub et al. procedure with the number in the Petrie procedure left the latter with substantially higher reliabilities. These results suggest that the 45-minute rest period may be critical to the full manifestation of the kinesthetic aftereffect, but a direct test of its causal role regarding differential reliabilities has not been undertaken.

Baker et al. (1976) continued the search for personality correlates of kinesthetic aftereffect begun by Petrie (1967). Inferences from her augmenter-reducer conception are that augmenters (their postinduction estimates smaller/narrower than their preinduction estimates) are overloaded with stimulation and hence motivated to avoid any more, whereas reducers (their postindduction estimated larger/wider than their preinduction estimates) are stimulus deprived and hence seek more stimulation. Supporting these inferences is empirical evidence (Petrie et al. 1958) indicating that reducers (relative to augmenters) are more tolerant of pain, whereas augmenters (relative to reducers) are more tolerant of sensory deprivation.

Baker et al. (1976), arguing that the first-session kinesthetic aftereffect was reliable and could potentially predict theoretically-relevant personality traits and behavioral dispositions, reanalyzed the earlier Weintraub et al. (1973) study. A 25-item scale was reduced to 18 items and an index was derived with positive scores reflecting the reducing end of the augmenting-reducing dimension. Some of the items in the index concerned responses to external stimulation (e.g., fear of an injection, lively parties, lengthy isolation). Other items concerned seeking or avoiding external stimulation (e.g., coffee and alcohol consumption, sports participation, smoking, friendship formation).

The kinesthetic-aftereffect scores for the first session were significantly related to the index ($r = -.36$, $p < .02$), as predicted, but the scores for the six subsequent sessions were not. Neither of the components of the kinesthetic-aftereffect score—preinduction and postinduction—were related to the index for any session. However, it is noteworthy that scores for subsequent sessions, made up from the preinduction score for the first session and the postinduction score for the subsequent session, were consistently related to the index. Baker et al. (1976) ended their article on a note of confidence, convinced that the kinesthetic-aftereffect task elicits personality differences in an augmenter-reducer dimension relevant to the manner in which external stimulation is sought and handled.

Nevertheless, in the very next year, an article by Herzog and Weintraub (1977) reported that an exact replication of the Baker et al. (1976) study found no link between the kinesthetic aftereffect and the personality behavior index. Herzog and Weintraub did, however, acknowledge the emergence of a reliable augmenter-reducer dimension. Disinclined to let the issue rest with so sharp a divergence from the Baker et al. study findings about the association between the kinesthetic aftereffect and the index, Herzog and Weintraub (1982) undertook a further replication. A slight procedural modification was introduced. Having failed to replicate the Baker et al. study with a wide inducing block, they chose to try a narrow inducing block for the first kinesthetic-aftereffect session and alternated the wide and narrow inducing blocks across subsequent sessions. Again, the results were negative, with the authors concluding that "we are unable to document any relationship between induction measures derived by the traditional kinesthetic-aftereffect procedure and questionnaire-derived personality measures" (Herzog and Weintraub 1982, p. 737).

Refusing to abandon the topic, Herzog et al. (1985) judged a final effort worthwhile if optimal procedures, identified in previous research, were applied. Accordingly, they employed the Petrie (1967) procedure (with the 45-minute initial rest period) that had previously generated exceptionally high reliabilities. They also selected the wide inducing block that had almost always been used whenever significant correlations with personality variables were obtained. In addition to the standard difference-score, Herzog et al. computed a residual change-score, "the deviation from the linear regression of post-induction scores on pre-induction scores" (p. 1342). In regard to the personality-behavior variables, a battery of measures was employed: a new 45-item questionnaire with two factor scales; the personality-behavior index used by Baker et al. (1976) and Herzog and Weintraub (1977, 1982); and several behavioral measures. Only those personality-behavior variables that had satisfactory internal-consistency reliability and at least two significant correlations with each other were retained for further analyses. All of these measurement and methodological precautions paid off in the demonstration that the kinesthetic aftereffect is indeed related to personality and behavior. Reducers (especially women) were significantly higher on the factor subscale Need for Sensory Stimulation, whose items have much in common with those on Zuckerman's (1994) sensation-seeking instruments. Consistent with earlier findings by Petrie et al. (1958) and Ryan and Foster (1967), reducers claimed to be more tolerant of cold temperatures and pain.

In sum, Herzog et al. (1985) have shown that the Petrie (1967) induction procedure generates reliable kinesthetic-aftereffect scores that correlate in the theoretically expected direction with reliable measures of personality and behavior. It is testimony to the importance of reliability when attempting to demonstrate the construct validity of a conceptually derived variable. However, a major disadvantage of the Petrie procedure must be acknowledged—an hour of individual administration—that is likely to limit the incentive of investigators to pursue further research with the procedure. It is hardly surprising then that research on the personality implications of the kinesthetic aftereffect essentially ended with the Herzog et al. investigation.

Virtually all of the research on the kinesthetic aftereffect-personality relationship has been interindividual (trait based). Baker et al. (1979) can be credited with one of the very few studies to explore intraindividual (state-based) variation. Baker et al. sought to determine whether the menstrual cycle influences kinesthetic-aftereffect. On the basis of evidence that maximal pain occurs at the beginning and end of the cycle, Baker et al. predicted greater kinesthetic aftereffect reduction (a larger aftereffect), "damping down of subjective intensity of incoming stimulation" (p. 236) at those points in the cycle and hence a curvilinear relationship between the kinesthetic aftereffect and locus in the menstrual cycle. Employing three samples of college-age women, quadratic-trend analysis yielded a significant curvilinear effect. The effect remained statistically significant when controlling for possible confounding variables—tiredness, oral contraceptive use, use of drugs or medication. Untested is the possibility of social-expectancy effects, participants at or near menses doing more poorly on the kinesthetic-aftereffect task simply because they believe women do poorly then. But, as Baker et al. observed, it is difficult to conceive of such effects in so unfamiliar a domain as the kinesthetic aftereffect.

Did personality research related to the kinesthetic aftereffect disappear from the psychological scene following the definitive Herzog et al. (1985) study? Not entirely, for the personality questionnaire used to validate the kinesthetic aftereffect became the primary instrument for assessing the augmenter-reducer dimension initially made operational in the kinesthetic-aftereffect laboratory tasks. A prime example of this change is represented by the Larsen and Zarate (1991) study. The 45-item questionnaire developed by Herzog et al. shifted from dependent to independent variable and was used to compare reducers' and augmenters' reactions to taking part in a boring and monotonous task. Compared to augmenters, reducers described the task as more aversive and were less likely to repeat it. Further, reducers, relative to augmenters, exhibited more novelty seeking and sensation seeking in their day-to-day activities.

Despite its promise, the augmenter-reducer construct seems to have vanished from the contemporary personality scene. Thus, it is absent from the index of the latest edition of the *Handbook of Personality* (John et al. 2008). The disappearance readily lends itself to speculation and a possible explanation. When there is a senior, prestigious psychologist advancing a construct whose predictions are highly similar to a construct advanced by younger psychologists of lesser reputation, the former's construct is likely to win out. Consider the theory of extraversion-introversion

developed by Eysenck (e.g., Eysenck and Eysenck 1985). Under quiet and calm conditions, extraverts and introverts are presumed to be equally aroused. But when external stimulation becomes excessive—bright lights, loud noises, crowds—introverts choose to withdraw so as to return to what for them is optimal stimulation. Extraverts, by contrast, need that kind of excitement to arrive at what for them is optimal stimulation. It is readily apparent that the more recent introversion-extraversion construct is virtually indistinguishable from the earlier augmenter-reducer construct. Given the central role of the introversion-extraversion concept in personality-trait theory and the similarity in the two constructs' theoretical links with personality, it is no surprise that the augmenter-reducer construct has faded away.

14.5 Conclusion

The conclusions about ETS research in cognitive, personality, and social psychology in the companion chapter (Stricker, Chap. 13, this volume) apply equally to the work described here: the remarkable breadth of the research in terms of the span of topics addressed (kinesthetic aftereffect to risk taking), the scope of the methods used (experiments, correlational studies, multivariate analyses), and the range of populations studied (young children, college and graduate students in the United States and Europe, the general public); its major impact on the field of psychology; and its focus on basic research.

Another conclusion can also be drawn from this work: ETS was a major center for research in creativity, cognitive styles, and risk taking in the 1960s and 1970s, a likely product of the fortunate juxtaposition of a supportive institutional environment, ample internal and external funding, and a talented and dedicated research staff.

Acknowledgments Thanks are due to Rachel Adler, Irene Kostin, and Nick Telepak for their assistance in retrieving publications, reports, and archival material, and to Randy Bennett, Jeremy Burrus, James Kaufman, and Donald Powers for reviewing a draft of this chapter.

References

Amabile, T. M. (1982). Social psychology of creativity: A consensual assessment technique. *Journal of Personality and Social Psychology, 43*, 997–1013. https://doi.org/10.1037/0022-3514.43.5.997

Atkinson, J. W. (1957). Motivational determinants of risk-taking behavior. *Psychological Review, 64*, 359–372. https://doi.org/10.1037/h0043445

Baer, J. (2011). How divergent thinking tests mislead us: Are the Torrance tests still relevant in the 21st century? The Division 10 debate. *Psychology of Aesthetics, Creativity, and the Arts, 5*, 309–313. https://doi.org/10.1037/a0025210

Baird, L. L. (1979). *Development of an inventory of documented accomplishments for graduate admissions* (GRE Board Research Report No. 77-3R). Princeton: Educational Testing Service.

Baird, L. L., & Knapp, J. E. (1981). *The Inventory of Documented Accomplishments for Graduate Admissions: Results of a field trial study of its reliability, short-term correlates, and evaluation* (GRE Board Research Report No. 78-3R). Princeton: Educational Testing Service. https://doi.org/10.1002/j.2333-8504.1981.tb01253.x

Baker, A. H., Mishara, B. L., Kostin, I. W., & Parker, L. (1976). Kinesthetic aftereffect and personality: A case study of issues involved in construct validation. *Journal of Personality and Social Psychology, 34*, 1–13. https://doi.org/10.1037/0022-3514.34.1.1

Baker, A. H., Mishara, B. L., Parker, L., & Kostin, I. W. (1978). When "reliability" fails, must a measure be discarded?—The case of kinesthetic aftereffect. *Journal of Research in Personality, 12*, 262–273. https://doi.org/10.1016/0092-6566(78)90053-3

Baker, A. H., Mishara, B. L., Kostin, I. W., & Parker, L. (1979). Menstrual cycle affects kinesthetic aftereffect, an index of personality and perceptual style. *Journal of Personality and Social Psychology, 37*, 234–246. https://doi.org/10.1037/0022-3514.37.2.234

Baker, A. H., Mishara, B. L., & Kostin, I. W. (1986). Kinesthetic aftereffect: One phenomenon or two? *Perception and Psychophysics, 39*, 255–260. https://doi.org/10.3758/BF03204932

Barron, F. (1953). Complexity-simplicity as a personality dimension. *Journal of Abnormal and Social Psychology, 48*, 163–172. https://doi.org/10.1037/h0054907

Bem, D. J., Wallach, M. A., & Kogan, N. (1965). Group decision making under risk of aversive consequences. *Journal of Personality and Social Psychology, 1*, 453–460. https://doi.org/10.1037/h0021803

Bennett, R. E., & Rock, D. A. (1995). Generalizability, validity, and examinee perceptions of a computer-delivered formulating-hypotheses test. *Journal of Educational Measurement, 32*, 19–36. https://doi.org/10.1111/j.1745-3984.1995.tb00454.x

Bromiley, P., & Curley, S. P. (1992). Individual differences in risk taking. In J. F. Yates (Ed.), *Risk-taking behavior* (pp. 87–132). New York: Wiley.

Brown, R. (1965). *Social psychology*. New York: Free Press.

Cronbach, L. J., & Snow, R. E. (1977). *Aptitudes and instructional methods*. New York: Irvington.

Crowne, D. P., & Marlowe, D. (1964). *The approval motive*. New York: Wiley.

Davis, J. K. (1991). Educational implications of field dependence-independence. In S. Wapner & J. Demick (Eds.), *Field dependence-independence: Cognitive style across the life span* (pp. 149–175). Hillsdale: Erlbaum.

Davis, J. H., Kameda, T., & Stasson, M. F. (1992). Group risk taking: Selected topics. In J. F. Yates (Ed.), *Risk-taking behavior* (pp. 163–200). New York: Wiley.

Edwards, W. (1961). Behavioral decision theory. *Annual Review of Psychology, 12*, 473–498. https://doi.org/10.1146/annurev.ps.12.020161.002353

Enright, M. K., Rock, D. A., & Bennett, R. E. (1998). Improving measurement for graduate admissions. *Journal of Educational Measurement, 35*, 250–267. https://doi.org/10.1111/j.1745-3984.1998.tb00538.x

Eysenck, H. J., & Eysenck, M. W. (1985). *Personality and individual differences: A natural science approach*. New York: Plenum. https://doi.org/10.1007/978-1-4613-2413-3

Figner, B., & Weber, E. U. (2011). Who takes risks when and why: Determinants of risk taking. *Current Directions in Psychological Science, 20*, 211–216. https://doi.org/10.1177/0963721411415790

Flanagan, J. C. (1949). *Critical requirements for research personnel—A study of observed behaviors of personnel in research laboratories*. Pittsburgh: American Institute for Research.

Frederiksen, N. (1959). *Development of the test "Formulating Hypotheses": A progress report* (ONR Technical Report, Contract Nonr-2338[00]). Princeton: Educational Testing Service.

Frederiksen, N., & Evans, F. R. (1974). Effects of models of creative performance on ability to formulate hypotheses. *Journal of Educational Psychology, 66*, 67–82. https://doi.org/10.1037/h0035808

Frederiksen, N., & Ward, W. C. (1978). Measures for the study of creativity in scientific problem-solving. *Applied Psychological Measurement, 2*, 1–24. https://doi.org/10.1177/014662167800200101

French, J. W., Ekstrom, R. B., & Price, L. A. (1963). *Manual for Kit of Reference Tests for Cognitive Factors*. Princeton: Educational Testing Service.

Gardner, R. W., Holzman, P. S., Klein, G. S., Linton, H. B., & Spence, D. P. (1959). Cognitive control: A study of individual consistencies in cognitive behavior. *Psychological Issues, 1*(4, Whole No. 4).

Gardner, R. W., Jackson, D. N., & Messick, S. (1960). Personality organization in cognitive controls and intellectual abilities. *Psychological Issues, 2*(4, Whole No. 8). https://doi.org/10.1037/11215-000

Getzels, J. W., & Jackson, P. W. (1962). *Creativity and intelligence: Explorations with gifted students*. New York: Wiley.

Glover, J. A., Ronning, R. R., & Reynolds, C. R. (Eds.). (1989). *Handbook of creativity*. New York: Plenum. https://doi.org/10.1007/978-1-4757-5356-1

Goodenough, D. R. (1976). The role of individual differences in field dependence as a factor in learning and memory. *Psychological Bulletin, 83*, 675–694. https://doi.org/10.1037/0033-2909.83.4.675

Goodenough, D. R. (1981). An entry in the great frame-tilt judging contest. *Perceptual and Motor Skills, 52*, 43–46. https://doi.org/10.2466/pms.1981.52.1.43

Goodenough, D. R. (1986). History of the field dependence construct. In M. Bertini, L. Pizzamiglio, & S. Wapner (Eds.), *Field dependence in psychological theory, research, and application* (pp. 5–13). Hillsdale: Erlbaum.

Goodenough, D. R., Oltman, P. K., & Cox, P. W. (1987). The nature of individual differences in field dependence. *Journal of Research in Personality, 21*, 81–99. https://doi.org/10.1016/0092-6566(87)90028-6

Goodenough, D. R., Oltman, P. K., Snow, D., Cox, P. W., & Markowitz, D. (1991). Field dependence-independence and embedded figures performance. In S. Wapner & J. Demick (Eds.), *Field dependence-independence: Cognitive style across the life span* (pp. 131–148). Hillsdale: Erlbaum.

Guilford, J. P. (1967). *The nature of human intelligence*. New York: McGraw-Hill.

Guilford, J. P., & Christensen, P. R. (1956). *A factor-analytic study of verbal fluency* (Psychological Laboratory Report No. 17). Los Angeles: University of Southern California.

Hennessey, B. A., & Amabile, T. M. (2010). Creativity. *Annual Review of Psychology, 61*, 569–598. https://doi.org/10.1146/annurev.psych.093008.100416

Hermann, M. G., & Kogan, N. (1968). Negotiation in leader and delegate groups. *Journal of Conflict Resolution, 12*, 322–344. https://doi.org/10.1177/002200276801200304

Herzog, T. R., & Weintraub, D. J. (1977). Preserving the kinesthetic aftereffect: Alternating inducing blocks day by day. *American Journal of Psychology, 90*, 461–474. https://doi.org/10.2307/1421876

Herzog, T. R., & Weintraub, D. J. (1982). Roundup time at personality ranch: Branding the elusive augmenters and reducers. *Journal of Personality and Social Psychology, 42*, 729–737. https://doi.org/10.1037/0022-3514.42.4.729

Herzog, T. R., Williams, D. M., & Weintraub, D. J. (1985). Meanwhile, back at personality ranch: The augmenters and reducers ride again. *Journal of Personality and Social Psychology, 48*, 1342–1352. https://doi.org/10.1037/0022-3514.48.5.1342

Holland, J. L. (1961). Creative and academic performance among talented adolescents. *Journal of Educational Psychology, 52*, 136–147. https://doi.org/10.1037/h0044058

Hovecar, D., & Bachelor, P. (1989). A taxonomy and critique of measurements used in the study of creativity. In J. A. Glover, R. R. Ronning, & C. R. Reynolds (Eds.), *Handbook of creativity* (pp. 53–75). New York: Plenum.

Irwin, F. W., & Smith, W. A. S. (1957). Value, cost, and information as determiners of decision. *Journal of Experimental Psychology, 54*, 229–232. https://doi.org/10.1037/h0049137

John, O. P., Robins, R. W., & Pervin, L. A. (Eds.). (2008). *Handbook of personality: Theory and research* (3rd ed.). New York: Guilford Press.

Kagan, J., & Kogan, N. (1970). Individual variation in cognitive processes. In P. H. Mussen (Ed.), *Carmichael's manual of child psychology* (Vol. 1, 3rd ed., pp. 1273–1365). New York: Wiley.

Kaufman, J. C., & Sternberg, R. J. (Eds.). (2006). *The international handbook of creativity*. New York: Cambridge University Press. https://doi.org/10.1017/CBO9780511818240

Kaufman, J. C., & Sternberg, R. J. (2010a). Preface. In J. C. Kaufman & R. J. Sternberg (Eds.). *The Cambridge handbook of creativity* (pp. xiii–xv). New York: Cambridge University Press. https://doi.org/10.1017/CBO9780511763205.001

Kaufman, J. C., & Sternberg, R. J. (Eds.). (2010b). *The Cambridge handbook of creativity*. New York: Cambridge University Press.

Kim, K. H. (2011). The APA 2009 Division 10 debate: Are the Torrance Tests of Creative Thinking still relevant in the 21st century? *Psychology of Aesthetics, Creativity, and the Arts, 5*, 302–308. https://doi.org/10.1037/a0021917

Kirton, M. J. (Ed.). (1989). *Adaptors and innovators*. London: Routledge.

Klein, S. P., & Skager, R. W. (1967). "Spontaneity vs deliberateness" as a dimension of esthetic judgment. *Perceptual and Motor Skills, 25*, 161–168. https://doi.org/10.2466/pms.1967.25.1.161

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. (1971). Educational implications of cognitive styles. In G. Lesser (Ed.), *Psychology and educational practice* (pp. 242–292). Glencoe: Scott Foresman.

Kogan, N. (1976). *Cognitive styles in infancy and early childhood*. Hillsdale: Erlbaum.

Kogan, N. (1980). A style of life, a life of style [Review of the book *Cognitive styles in personal and cultural adaptation* by H. A. Witkin]. *Contemporary Psychology, 25*, 595–598. https://doi.org/10.1037/018043

Kogan, N. (1983). Stylistic variation in childhood and adolescence: Creativity, metaphor, and cognitive styles. In J. Flavell & E. Markman (Eds.), *Handbook of child psychology: Vol. 3. Cognitive development* (4th ed., pp. 630–706). New York: Wiley.

Kogan, N. (1994). Cognitive styles. In R. J. Sternberg (Ed.), *Encyclopedia of human intelligence* (pp. 266–273). New York: Macmillan.

Kogan, N. (2008). Commentary: Divergent-thinking research and the Zeitgeist. *Psychology of Aesthetics, Creativity, and the Arts, 2*, 10–102. https://doi.org/10.1037/1931-3896.2.2.100

Kogan, N., & Block, J. (1991). Field dependence-independence from early childhood through adolescence: Personality and socialization aspects. In S. Wapner & J. Demick (Eds.), *Field dependence-independence: Cognitive style across the life span* (pp. 177–207). Hillsdale: Erlbaum.

Kogan, N., & Carlson, J. (1969). Difficulty of problems attempted under conditions of competition and group consensus. *Journal of Educational Psychology, 60*, 155–167. https://doi.org/10.1037/h0027555

Kogan, N., & Doise, W. (1969). Effects of anticipated delegate status on level of risk taking in small decision-making groups. *Acta Psychologica, 29*, 228–243. https://doi.org/10.1016/0001-6918(69)90017-1

Kogan, N., & Morgan, F. T. (1969). Task and motivational influences on the assessment of creative and intellective ability in children. *Genetic Psychology Monographs, 80*, 91–127.

Kogan, N., & Pankove, E. (1972). Creative ability over a five-year span. *Child Development, 43*, 427–442. https://doi.org/10.2307/1127546

Kogan, N., & Pankove, E. (1974). Long-term predictive validity of divergent-thinking tests: Some negative evidence. *Journal of Educational Psychology, 66*, 802–810. https://doi.org/10.1037/h0021521

Kogan, N., & Saarni, C. (1990). Cognitive styles in children: Some evolving trends. In O. N. Saracho (Ed.), *Cognitive style and early education* (pp. 3–31). New York: Gordon and Breach.
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
Kogan, N., & Wallach, M. A. (1967c). Risk taking as a function of the situation, the person, and the group. *New Directions in Psychology, 3*, 111–289.
Kogan, N., & Wallach, M. A. (1967d). Risky-shift phenomenon in small decision-making groups: A test of the information-exchange hypothesis. *Journal of Experimental Social Psychology, 3*, 75–84.
Kogan, N., Lamm, H., & Trommsdorff, G. (1972). Negotiation constraints in the risk-taking domain: Effects of being observed by partners of higher or lower status. *Journal of Personality and Social Psychology, 23*, 143–156. https://doi.org/10.1037/h0033035
Köhler, W., & Dinnerstein, D. (1947). Figural after-effects in kinesthesis. In *Miscellanea psychologica Albert Michotte* (pp. 196–220). Paris, France: Librairie Philosophique.
Kolb, D. A. (1976). *Learning Style Inventory: Technical manual*. Englewood Cliffs: Prentice Hall.
Kozhevnikov, M. (2007). Cognitive styles in the context of modern psychology: Toward an integrative framework of cognitive style. *Psychological Bulletin, 133*, 464–481. https://doi.org/10.1037/0033-2909.133.3.464
Lamm, H., & Kogan, N. (1970). Risk taking in the context of intergroup negotiation. *Journal of Experimental Social Psychology, 6*, 351–363. https://doi.org/10.1016/0022-1031(70)90069-7
Lamm, H., Trommsdorff, G., & Kogan, N. (1970). Pessimism-optimism and risk taking in individual and group contexts. *Journal of Personality and Social Psychology, 15*, 366–374. https://doi.org/10.1037/h0029602
Larsen, R. J., & Zarate, M. A. (1991). Extending reducer/augmenter theory into the emotion domain: The role of affect in regulating stimulation level. *Personality and Individual Differences, 12*, 713–723. https://doi.org/10.1016/0191-8869(91)90227-3
Le Bon, G. (1960). *The crowd: A study of the popular mind*. New York: Viking. (Original work published 1895).
Linn, M. C., & Kyllonen, P. (1981). The field dependence-independence construct: Some, one, or none. *Journal of Educational Psychology, 73*, 261–273. https://doi.org/10.1037/0022-0663.73.2.261
MacKinnon, D. W. (1962). The nature and nurture of creative talent. *American Psychologist, 17*, 484–495. https://doi.org/10.1037/h0046541
Mednick, S. A. (1962). The associative basis of the creative process. *Psychological Review, 69*, 220–232. https://doi.org/10.1037/h0048850
Mednick, S. A., & Mednick, M. T. (1962). *Examiner's manual, Remote Associates Test*. Rolling Meadows: Riverside.
Messick, S. (1970). The criterion problem in the evaluation of instruction: Assessing possible, not just intended, outcomes. In M. C. Wittrock & D. W. Wiley (Eds.), *The evaluation of instruction: Issues and problems* (pp. 183–202). New York: Holt Rinehart and Winston.
Messick, S. (1976). Personality consistencies in cognition and creativity. In S. Messick (Ed.), *Individuality in learning* (pp. 4–22). San Francisco: Jossey-Bass.
Messick, S. (1984). The nature of cognitive styles: Problems and promise in educational practice. *Educational Psychologist, 19*, 59–74. https://doi.org/10.1080/00461528409529283
Messick, S. (1987). Structural relationships across cognition, personality, and style. In R. E. Snow & M. J. Farr (Eds.), *Aptitude, learning, and instruction: Vol. 3. Conative and affective process analysis* (pp. 35–75). Hillsdale: Erlbaum.
Messick, S. (1994). The matter of style: Manifestations of personality in cognition, learning, and teaching. *Educational Psychologist, 29*, 121–136. https://doi.org/10.1207/s15326985ep2903_2

Messick, S. (1996). Bridging cognition and personality in education: The role of style in performance and development. *European Journal of Personality, 10*, 353–376. https://doi.org/10.1002/(SICI)1099-0984(199612)10:5<353::AID-PER268>3.0.CO;2-G

Messick, S., & Damarin, F. (1964). Cognitive styles and memory for faces. *Journal of Abnormal and Social Psychology, 69*, 313–318. https://doi.org/10.1037/h0044501

Messick, S., & Fritzky, F. J. (1963). Dimensions of analytic attitude in cognition and personality. *Journal of Personality, 31*, 346–370. https://doi.org/10.1111/j.1467-6494.1963.tb01304.x

Messick, S., & Kogan, N. (1963). Differentiation and compartmentalization in object-sorting measures of categorizing style. *Perceptual and Motor Skills, 16*, 47–51. https://doi.org/10.2466/pms.1963.16.1.47

Messick, S., & Kogan, N. (1965). Category width and quantitative aptitude. *Perceptual and Motor Skills, 20*, 493–497. https://doi.org/10.2466/pms.1965.20.2.493

Messick, S., & Kogan, N. (1966). Personality consistencies in judgment: Dimensions of role constructs. *Multivariate Behavioral Research, 1*, 165–175. https://doi.org/10.1207/s15327906mbr0102_3

Mishara, B. L., & Baker, A. H. (1978). Kinesthetic aftereffect scores are reliable. *Applied Psychological Measurement, 2*, 239–247. https://doi.org/10.1177/014662167800200206

Moscovici, S., & Doise, W. (1994). *Conflict and consensus*. Thousand Oaks: Sage.

Myers, D. G., & Lamm, H. (1976). The group polarization phenomenon. *Psychological Bulletin, 83*, 602–627. https://doi.org/10.1037/0033-2909.83.4.602

Pankove, E., & Kogan, N. (1968). Creative ability and risk taking in elementary school children. *Journal of Personality, 36*, 420–439. https://doi.org/10.1111/j.1467-6494.1968.tb01483.x

Paunonen, S. V., & Jackson, D. N. (1985). Idiographic measurement strategies for personality and prediction: Some unredeemed promissory notes. *Psychological Review, 92*, 486–511. https://doi.org/10.1037/0033-295X.92.4.486

Petrie, A. (1967). *Individuality in pain and suffering*. Chicago: University of Chicago Press.

Petrie, A., Collins, W., & Solomon, P. (1958). Pain sensitivity, sensory deprivation, and susceptibility to satiation. *Science, 128*, 1431–1433. https://doi.org/10.1126/science.128.3336.1431-a

Plucker, J. A. (1990). Reanalyses of student responses to creativity checklists: Evidence of content generality. *Journal of Creative Behavior, 33*, 126–137. https://doi.org/10.1002/j.2162-6057.1999.tb01042.x

Plucker, J. A. (1999). Is there proof in the pudding? Reanalyses of Torrance's (1958 to present) longitudinal data. *Creativity Research Journal, 12*, 103–114. https://doi.org/10.1207/s15326934crj1202_3

Plucker, J. A., & Makel, M. C. (2010). Assessment of creativity. In J. C. Kaufman & R. J. Sternberg (Eds.), *The Cambridge handbook of creativity* (pp. 48–79). New York: Cambridge University Press. https://doi.org/10.1017/CBO9780511763205.005

Pruitt, D. G. (1962). Pattern and level of risk in gambling decisions. *Psychological Review, 69*, 187–201. https://doi.org/10.1037/h0040607

Ryan, E. D., & Foster, R. (1967). Athletic participation and perceptual reduction and augmentation. *Journal of Personality and Social Psychology, 6*, 472–476. https://doi.org/10.1037/h0021226

Saunders, D. R. (1956). Moderator variables in prediction. *Educational and Psychological Measurement, 16*, 209–222. https://doi.org/10.1177/001316445601600205

Silvia, P. J. (2008). Creativity and intelligence revisited: A latent variable analysis of Wallach and Kogan (1965). *Creativity Research Journal, 20*, 34–39. https://doi.org/10.1080/10400410701841807

Silvia, P. J., Winterstein, B. P., Willse, J. T., Barona, C. M., Cram, J. T., Hess, K. I., et al. (2008). Assessing creativity with divergent-thinking tasks: Exploring the reliability and validity of new subjective scoring methods. *Psychology of Aesthetics, Creativity, and the Arts, 2*, 68–85. https://doi.org/10.1037/1931-3896.2.2.68

Simon, H. A. (1957). *Administrative behavior*. New York: Macmillan.

Skager, R. W., Schultz, C. B., & Klein, S. P. (1965). Quality and quantity of accomplishments as measures of creativity. *Journal of Educational Psychology, 56*, 31–39. https://doi.org/10.1037/h0021901

Skager, R. W., Schultz, C. B., & Klein, S. P. (1966a). Points of view about preference as tools in the analysis of creative products. *Perceptual and Motor Skills, 22*, 83–94. https://doi.org/10.2466/pms.1966.22.1.83

Skager, R. W., Schultz, C. B., & Klein, S. P. (1966b). The multidimensional scaling of a set of artistic drawings: Perceived structure and scale correlates. *Multivariate Behavioral Research, 1*, 425–436. https://doi.org/10.1207/s15327906mbr0104_2

Slovic, P. (1962). Convergent validation of risk taking measures. *Journal of Abnormal and Social Psychology, 65*, 68–71. https://doi.org/10.1037/h0048048

Spence, J. T., & Spence, K. W. (1966). The motivational components of manifest anxiety: Drive and drive stimuli. In C. D. Spielberger (Ed.), *Anxiety and behavior* (pp. 291–326). San Diego: Academic Press. https://doi.org/10.1016/B978-1-4832-3131-0.50017-2

Sternberg, R. J. (1997). *Thinking styles*. New York: Cambridge University Press. https://doi.org/10.1017/CBO9780511584152

Sternberg, R. J. (Ed.). (1999). *Handbook of creativity*. New York: Cambridge University Press.

Stricker, L. J., & Rock, D. A. (2001). Measuring accomplishments: Pseudoipsativity, quantity versus quality, and dimensionality. *Personality and Individual Differences, 31*, 103–115. https://doi.org/10.1016/S0191-8869(00)00114-8

Taylor, C., & Barron, F. (Eds.). (1963). *Scientific creativity*. New York: Wiley.

Teger, A. I., & Kogan, N. (1975). Decision making for others under reciprocal and non-reciprocal conditions. *British Journal of Social and Clinical Psychology, 14*, 215–222. https://doi.org/10.1111/j.2044-8260.1975.tb00174.x

Thibaut, J. W., & Kelley, H. H. (1959). *The social psychology of groups*. New York: Wiley.

Thomas, K., & Chan, J. (Eds.). (2013). *Handbook of research on creativity*. Cheltenham: Elgar.

Thorndike, R. L. (1966). Some methodological issues in the study of creativity. In A. Anastasi (Ed.), *Testing problems in perspective—Twenty-fifth anniversary volume of topical readings from the Invitational Conference on Testing Problems* (pp. 436–448). Washington, DC: American Council on Education.

Torrance, E. P. (1962). *Guiding creative talent*. Englewood Cliffs: Prentice-Hall. https://doi.org/10.1037/13134-000

Torrance, E. P. (1974). *Norms-technical manual, Torrance Tests of Creative Thinking*. Bensenville: Scholastic Testing Service.

Vernon, P. E. (1972). The distinctiveness of field independence. *Journal of Personality, 40*, 366–391. https://doi.org/10.1111/j.1467-6494.1972.tb00068.x

Wallach, M. A. (1970). Creativity. In P. H. Mussen (Ed.), *Carmichael's manual of child psychology* (Vol. 1, 3rd ed., pp. 1211–1272). New York: Wiley.

Wallach, M. A. (1971). *The intelligence/creativity distinction*. Morristown: General Learning Press.

Wallach, M. (1976). Tests tell us little about talent. *American Scientist, 64*, 57–63.

Wallach, M. A., & Kogan, N. (1961). Aspects of judgment and decision making: Interrelationships and changes with age. *Behavioral Science, 6*, 23–36. https://doi.org/10.1002/bs.3830060104

Wallach, M. A., & Kogan, N. (1965a). *Modes of thinking in young children—A study of the creativity-intelligence distinction*. New York: Holt Rinehart and Winston.

Wallach, M. A., & Kogan, N. (1965b). The roles of information, discussion, and consensus in group risk taking. *Journal of Experimental Social Psychology, 1*, 1–19. https://doi.org/10.1016/0022-1031(65)90034-X

Wallach, M. A., & Wing Jr., C. W. (1969). *The talented student: A validation of the creativity/intelligence distinction*. New York: Holt Rinehart and Winston.

Wallach, M. A., Kogan, N., & Bem, D. J. (1962). Group influence on individual risk taking. *Journal of Abnormal and Social Psychology, 65*, 75–86. https://doi.org/10.1037/h0044376

Wallach, M. A., Kogan, N., & Bem, D. J. (1964). Diffusion of responsibility and level of risk taking in groups. *Journal of Abnormal and Social Psychology, 68*, 263–274. https://doi.org/10.1037/h0042190

Wallach, M. A., Kogan, N., & Burt, R. B. (1965). Can group members recognize the effects of group discussion upon risk taking? *Journal of Experimental Social Psychology, 1*, 379–395. https://doi.org/10.1016/0022-1031(65)90016-8

Wallach, M. A., Kogan, N., & Burt, R. B. (1967). Group risk taking and field dependence-independence of group members. *Sociometry, 30*, 323–338.

Wallach, M. A., Kogan, N., & Burt, R. B. (1968). Are risk takers more persuasive than conservatives in group discussion? *Journal of Experimental Social Psychology, 4*, 76–88. https://doi.org/10.1016/0022-1031(68)90051-6

Wallbrown, F. H., & Huelsman Jr., C. B. (1975). The validity of the Wallach-Kogan creativity operations for the inner-city children in two areas of visual art. *Journal of Personality, 43*, 109–126. https://doi.org/10.1111/j.1467-6494.1975.tb00575.x

Wapner, S., & Demick, J. (Eds.). (1991). *Field dependence-independence: Cognitive style across the life span*. Hillsdale: Erlbaum.

Ward, W. C. (1969a). Creativity and environmental cues in nursery school children. *Developmental Psychology, 1*, 543–547. https://doi.org/10.1037/h0027977

Ward, W. C. (1969b). Rate and uniqueness in children's creative responding. *Child Development, 40*, 869–878. https://doi.org/10.2307/1127195

Ward, W. C. (1975). Convergent and divergent measurement of creativity in children. *Educational and Psychological Measurement, 35*, 87–95. https://doi.org/10.1177/001316447503500110

Ward, W. C., & Cox, P. W. (1974). A field study of nonverbal creativity. *Journal of Personality, 42*, 202–219. https://doi.org/10.1111/j.1467-6494.1974.tb00670.x

Ward, W. C., Kogan, N., & Pankove, E. (1972). Incentive effects in children's creativity. *Child Development, 43*, 669–676. https://doi.org/10.2307/1127565

Ward, W. C., Frederiksen, N., & Carlson, S. B. (1980). Construct validity of free-response and machine-scorable forms of a test. *Journal of Educational Measurement, 17*, 11–29. https://doi.org/10.1111/j.1745-3984.1980.tb00811.x

Weintraub, D. J., Green, G. S., & Herzog, T. R. (1973). Kinesthetic aftereffects day by day: Trends, task features, reliable individual differences. *American Journal of Psychology, 86*, 827–844. https://doi.org/10.2307/1422088

Witkin, H. A., & Berry, J. W. (1975). Psychological differentiation in cross-cultural perspective. *Journal of Cross-Cultural Psychology, 6*, 4–87. https://doi.org/10.1177/002202217500600102

Witkin, H. A., & Goodenough, D. R. (1977). Field dependence and interpersonal behavior. *Psychological Bulletin, 84*, 661–689. https://doi.org/10.1037/0033-2909.84.4.661

Witkin, H. A., & Goodenough, D. R. (1981). *Cognitive styles: Essence and origins*. New York: International Universities Press.

Witkin, H. A., Dyk, R. B., Faterson, H. F., Goodenough, D. R., & Karp, S. A. (1962). *Psychological differentiation*. New York: Wiley.

Witkin, H. A., Price-Williams, D., Bertini, M., Christiansen, B., Oltman, P. K., Ramirez, M., & van Meel, J. M. (1974). Social conformity and psychological differentiation. *International Journal of Psychology, 9*, 11–29. https://doi.org/10.1080/00207597408247089

Witkin, H. A., Moore, C. A., Oltman, P. K., Goodenough, D. R., Friedman, F., Owen, D. R., & Raskin, E. (1977a). Role of the field-dependent and field-independent cognitive styles in academic evolution: A longitudinal study. *Journal of Educational Psychology, 69,* 197–211. https://doi.org/10.1037/0022-0663.69.3.197

Witkin, H. A., Moore, C. A., Goodenough, D. R., & Cox, P. W. (1977b). Field-dependent and field-independent cognitive styles and their educational implications. *Review of Educational Research, 47*, 1–64. https://doi.org/10.3102/00346543047001001

Witkin, H. A., Goodenough, D. R., & Oltman, P. K. (1979). Psychological differentiation: Current status. *Journal of Personality and Social Psychology, 37*, 1127–1145. https://doi.org/10.1037/0022-3514.37.7.1127

Yates, J. F. (Ed.). (1992). *Risk-taking behavior*. New York: Wiley.
Zaleska, M., & Kogan, N. (1971). Level of risk selected by individuals and groups when deciding for self and for others. *Sociometry, 34*, 198–213. https://doi.org/10.2307/2786410
Zhang, L., & Sternberg, R. J. (2006). *The nature of intellectual styles*. Mahwah: Erlbaum.
Zhang, L., Sternberg, R. J., & Rayner, S. (Eds.). (2011). *Handbook of intellectual styles*. New York: Springer.
Zoccolotti, P., & Oltman, P. K. (1978). Field dependence and lateralization of verbal and configurational processing. *Cortex, 14*, 155–163. https://doi.org/10.1016/S0010-9452(78)80040-9
Zuckerman, M. (1994). *Behavioural expressions and biosocial bases of sensation seeking*. Cambridge: Cambridge University Press.

Chapter 15
Research on Developmental Psychology

Nathan Kogan, Lawrence J. Stricker, Michael Lewis, and Jeanne Brooks-Gunn

Developmental psychology was a major area of research at ETS from the late 1960s to the early 1990s, a natural extension of the work in cognitive, personality, and social psychology that had begun shortly after the organization's founding in 1947, consistent with Henry Chauncey's vision of investigating intellectual and personal qualities (see Stricker, Chap. 13, this volume). For a full understanding of these qualities, it is essential to know how they emerge and evolve. Hence the work in developmental psychology complemented the efforts already under way in other fields of psychology.

A great deal of the research in developmental psychology was conducted at ETS's Turnbull Hall in the Infant Laboratory, equipped with physiological recording equipment and observation rooms (e.g., Lewis 1974), and in a full-fledged Montessori school outfitted with video cameras (e.g., Copple et al. 1984). Hence, as Lewis (n.d.) recalled, the building "had sounds of infants crying and preschool children laughing" (p. 4). Other research was done in homes, schools, and hospitals, including a multisite longitudinal study of Head Start participants (e.g., Brooks-Gunn et al. 1989; Laosa 1984; Shipman 1972).

A handful of investigators directed most of the research, each carrying out a distinct program of extensive and influential work. This chapter covers research by Irving Sigel, on representational competence; Luis Laosa, on parental influences, migration, and measurement; Michael Lewis, on cognitive, personality, and social development of infants and young children; and Jeanne Brooks-Gunn, on cognitive,

N. Kogan • L.J. Stricker (✉)
Educational Testing Service, Princeton, NJ, USA
e-mail: lstricker@ets.org

M. Lewis
Rutgers Robert Wood Johnson Medical School, New Brunswick, NJ, USA

J. Brooks-Gunn
Columbia University, New York, NY, USA

R.E. Bennett, M. von Davier (eds.), *Advancing Human Assessment*,
Methodology of Educational Measurement and Assessment,
DOI 10.1007/978-3-319-58689-2_15

personality, and social development from infancy to adolescence. Other important research was conducted by Gordon Hale (e.g., Hale and Alderman 1978), on attention; Walter Emmerich (e.g., Emmerich 1968, 1982), on sex roles and personality development; and Nathan Kogan (e.g., Wallach and Kogan 1965) and William Ward (e.g., Ward 1968), on creativity. (The Kogan and Ward research is included in Kogan, Chap. 14, this volume.) In the present chapter, Kogan describes Sigel's research, and Stricker takes up Laosa's; Lewis and Brooks-Gunn discuss their own work.

15.1 Representational Competence and Psychological Distance

Representational competence was the focus of Sigel's research program. Roughly defined by Sigel and Saunders (1983), representational competence is the ability to transcend immediate stimulation and to remember relevant past events and project future possibilities. Also indicative of representational competence in preschoolers was the understanding of equivalence in symbol systems, whereby an object could be rendered three-dimensionally in pictorial form and in words.

The level of a child's representational competence was attributed in large part to parental beliefs and communicative behaviors and to family constellation (number of children and their birth order and spacing). Earlier research by Sigel and collaborators emphasized ethnicity and socioeconomic status (SES; see Kogan 1976). SES was retained in many of the ETS studies in addition to a contrast between typical children and those with communicative–language disabilities.

A conceptual model of the Sigel team's research approach is presented in a chapter by McGillicuddy-DeLisi et al. (1979): Mothers' and fathers' backgrounds determined their parental belief systems. Belief systems, in turn, influenced parental communication strategies, which then accounted for the child's level of cognitive development. It was a nonrecursive model, the child's developmental progress (relative to his or her age) feeding back to alter the parental belief systems. In terms of research design, then, parental background was the independent variable, parental belief systems and child-directed communicative behavior were mediating variables, and children's representational competence was the dependent variable. The full model was not implemented in every study, and other relevant variables were not included in the model. In most studies, family constellation (e.g., spacing and number of children), SES, the nature of the parent–child interaction task, the child's communicative status (with or without language disability), and the gender of the parent and child were shown to yield main or interaction effects on the child's representational competence.

In the view of Sigel and his associates, the critical component of parental teaching behavior was *distancing* (Sigel 1993). Parental teachings could reflect high- or low-level distancing. Thus, in a teaching context, asking the child to label an object was an example of low-level distancing, for the child's response was constrained to

a single option with no higher-thinking processes invoked in the answer. By contrast, asking the child to consider possible uses of an object was an example of high-level distancing, for the child was forced to go beyond the overt stimulus properties of the object to adopt new perspectives toward it. In brief, the concept of distancing, as reflected in parental teaching behavior, referred to the degree of constraint versus openness that the parent imposed on the child. Sigel's principal hypothesis was that higher-level distancing in child-directed communication by an adult would be associated with greater representational competence for that child. Correspondingly, low-level distancing by an adult would inhibit the development of a child's representational competence.

An additional feature of Sigel's research program concerned the nature of the task in the parent–child interaction. Two tasks were selected of a distinctively different character. For the storytelling task, age-appropriate edited versions of children's books were used, with parents instructed to go through a story as they typically would do at home. The other task required paper folding, with the parent required to teach the child to make a boat or a plane.

15.1.1 Influence of Parental Beliefs and Behavior on Representational Competence

Having outlined the conceptual underpinning of Sigel's research program along with the nature of the variables selected and the research designs employed, we can now proceed to describe specific studies in greater detail. We begin with a study of 120 families in which the target child was 4 years of age (McGillicuddy-DeLisi 1982; Sigel 1982). Family variables included SES (middle vs. working class) and single-child versus three-child families. For the three-child families, there was variation in the age discrepancy between the first and second sibling (more than 3 years apart vs. less than 3 years apart), with the restriction that siblings be of the same sex. Each mother and father performed the storytelling and paper-folding tasks with their 4-year-old child. Proper controls were employed for order of task presentations. A total of 800 parent and child observations were coded by six raters with satisfactory interrater reliability.

The presentation of the research was divided into two parts, corresponding to the portion of the analytic model under investigation. In the first part (McGillicuddy-DeLisi 1982), the influence of the demographic variables, SES and family constellation, on parental beliefs was examined, and in turn the influence of parental beliefs for their prediction of overt parental behaviors in a teaching situation was explored. The second part, the child's representational competence, was treated separately in the Sigel (1982) chapter. Note that the assessment of beliefs was focused exclusively on the parents' views of how a preschool child acquired concepts and abilities, hence making such beliefs relevant to the parental strategies employed in facilitating the child's performance in a teaching context.

Parental beliefs were assessed in an interview based on 12 vignettes involving a 4-year-old and a mother or father. The interviewer asked the parent whether the child in the vignette had the necessary concepts or abilities to handle the problem being posed. Further inquiry focused more generally on parents' views of how children acquire concepts and abilities. Analysis of these data yielded 26 parental belief variables that were reliably scored by three coders. ANOVA was then employed to determine the influence of SES, family constellation, gender of child, and gender of parent on each of the 26 belief variables. Beliefs were found to vary more as a function of SES and family constellation than of gender of parent or child. More specifically, parents of three children had views of child development that differed substantially from those of single-child parents. For the parents of three children, development involved attributes more internal to the child (e.g., reference to self-regulation and impulsivity) as opposed to greater emphasis on external attributes (e.g., direct instruction) in single-child parents. The results as a whole constituted an intriguing mosaic, but they were post hoc in the absence of predictions derived from a theoretical framework. Of course, the exploratory nature of such research reflected the dearth at that time of theoretical development in the study of child-directed parental beliefs and behaviors.

Consider next the observed relationships between parental beliefs and teaching behaviors. Having shown that SES and family constellation influenced parental beliefs, the question of interest was whether such beliefs provided useful information about parents' teaching behaviors beyond what might be predicted from SES and family constellation. To answer the question, stepwise regressions were carried out with SES and family constellation entered into the analysis first, followed by the belief variables. Separate regressions—four in all—were conducted for mothers' and fathers' performance on the storytelling and paper-folding tasks, the dependent variables.

Demonstration of belief effects on teaching behaviors would require that multiple correlations show significant increments in magnitude when beliefs were entered into the regression analysis. Such increments were observed in all four regressions, indicating that parents' beliefs about their children's competencies were predictive of the way they went about teaching their children on selected tasks. Noteworthy is the evidence that the significant beliefs varied across the two tasks and that this variation was greater for mothers than for fathers. In other words, mothers appeared to be more sensitive to the properties of the task facing the child, whereas fathers appeared to have internalized a set of beliefs generally applied to different kinds of tasks. Mothers would seem to have a more differentiated view of their children's competencies and hence were more attuned to the nature of the task than were fathers.

Thus far, we have considered the relations among family demographics, parental beliefs, and teaching strategies. The missing link, the child's cognitive performance, was examined in the Sigel (1982) chapter, where it was specifically related to parental teaching behaviors. The child's responses to the storytelling and paper-folding tasks were considered (e.g., extent of engagement and problem solutions), as was the child's performance on tasks independent of parental instructions. These latter

tasks included Piagetian conservation and imagery assessments and the Sigel Object Categorization Task (Sigel and Olmsted 1970). The major hypothesis was that the parents' uses of distancing strategies in their teaching behaviors would be associated with enhanced cognitive performances in their children—representational competence.

To address this hypothesis, stepwise regressions were analyzed. The results confirmed the basic hypothesis linking parental child-directed distancing to the child's representational competence. This general observation, however, conceals the specificity of the effects. Thus mothers and fathers employed different teaching strategies, and these strategies, in turn, varied across the storytelling and paper-folding tasks. Of special interest are those analyses in which the mothers' and fathers' teaching behaviors were entered into the same regression equation. Doing so in sequence often pointed to the complementarity of parental influences. In concrete terms, the multiple correlations sometimes demonstrated significant enhancements when both parents' teaching strategies entered into the analysis compared to the outcome for the parents considered separately. This result implied that the children could intellectually profit from the different, but complementary, teaching strategies of mothers and fathers.

15.1.2 Impact of a Communicative Disability

Sigel and McGillicuddy-DeLisi (1984) were able to recruit families who had a child with a communicative disability (CD), making it possible to compare such families with those where the child was not communicatively disabled (non-CD). It was possible to match the CD and non-CD children on SES, family size, gender, age, and birth order. Again, mothers' and fathers' distancing behaviors were examined in the task context of storytelling and paper folding.

In the case of the child's intellectual ability, assessed by the Wechsler Preschool and Primary School Scale of Intelligence (WPPSI; Wechsler 1949b), parental effects were largely confined to the CD sample. Low parental distancing strategies were tightly associated with lower WPPSI scores. Of course, we must allow for the possibility that the parent was adjusting his or her distancing level to the perceived cognitive ability of the child. In contrast, the child's representational competence, as defined by the assessments previously described in Sigel (1982), was linked with parental distancing behaviors in both CD and non-CD samples, with the magnitude of the relationship somewhat higher in the CD sample.

Of course, these associations could not address the causality question: The parent might be affecting the child or reacting to the child or, more likely, the influence was proceeding in both directions. Sigel and McGillicuddy-DeLisi (1984) argued that low-level distancing strategies by parents discouraged active thinking in the child; hence it was no surprise that such children did not perform well on representational tasks that required such thinking. They were optimistic about CD children, for high-level parental distancing seemed to encourage the kind of representational

thinking that could partially compensate for their communicative disabilities (Sigel 1986).

15.1.3 Belief-Behavior Connection

Working with a subsample of the non-CD families described in the previous section, Sigel (1992) plunged into the controversial issue of the linkage between an individual's beliefs and actual behavior instantiating those beliefs. He also developed a measure of behavioral intentions—a possible mediator of the belief–behavior connection. Although the focus was naturally on parental beliefs and behaviors, similar work in social psychology on the belief and behavior connection (e.g., Ajzen and Fishbein 1977), where major advances in theory and research had occurred, was not considered.

Three categories of variables were involved: (a) parents' beliefs about how children acquired knowledge in four distinct domains (physical, social, moral, and self); (b) the strategies that parents claimed they would use to facilitate the children's acquisition of knowledge in those domains; and (c) the behavioral strategies employed by the parents in a teaching context with their children. The first two categories were assessed with a series of vignettes. Thus, in the vignette for the physical domain, the child asks the parent how to use a yardstick to measure the capacity of their bathtub. The parents' view about how children learn about measurement constituted the belief measure; the parents' statements about how they would help their child learn about measurement constituted the self-referent strategy measure. For the third category, the parents taught their child how to tie knots, and the strategies employed in doing so were observed. Note that the knots task involved different content than was used in the vignettes.

Parental beliefs regarding children's learning were categorized as emphasizing cognitive processing (e.g., children figuring out things on their own) or direct instruction (e.g., children learning from being told things by adults). Parental intended teaching strategies were classified as distancing, rational authoritative (e.g., parent gives reasons with commands), or direct authoritative (e.g., parent offers statement or rule without rationale). Parental behavioral strategies were scored for high-level versus low-level distancing.

The three variable classes—parental beliefs, parental intended teaching strategies, and parental behavioral strategies—were intercorrelated. Substantial relationships were observed between parental beliefs about learning (cognitive processing vs. direct instruction) and the strategies the parent intended to employ. As anticipated, cognitive processing was associated with distancing strategies, and direct instruction was linked to authoritative strategies. Of course, both the beliefs and self-referent strategies were derived from the same vignettes used in the parental interview, suggesting the likely influence of method variance on the correlational outcomes. When the foregoing variables were related to the parents' behavioral strategies in teaching the knots task, the magnitude of the correlations dropped

precipitously, though the marginally significant correlations were in the predicted direction. Sigel (1992) attributed the correlational decline to variation across domains. Thus the belief–strategy linkages were not constant across physical, social, and moral problems. Aggregation across these domains could not be justified. Obviously, the shifting task content and context were also responsible for the absence of anticipated linkages. Conceivably, an analytic procedure in which parents' intended strategies were cast as mediators between their beliefs and their behavioral strategies would have yielded further enlightenment.

15.1.4 Collaborative Research

The large majority of Sigel's publications were either solely authored by him or coauthored with former or present members of his staff at ETS. A small number of papers, however, were coauthored with two other investigators, Anthony Pellegrini and Gene Brody, at the University of Georgia. These publications are of particular interest because they cast Sigel's research paradigm within a different theoretical framework, that of Vygotsky (1978), and they introduced a new independent variable into the paradigm, marital quality.

In the case of marital quality, Brody et al. (1986) raised the possibility that the quality of the marital relationship would influence mothers' and fathers' interactions with their elementary-school age children. More specifically, Brody et al., leaning on clinical reports, examined the assumption that marital distress would lead to compensatory behaviors by the parents when they interact with their children in a teaching context. Also under examination was the possibility that mothers and fathers would employ different teaching strategies when interacting with the children, with the nature of such differences possibly contingent on the levels of marital distress.

Again, storytelling and paper-folding tasks were used with the mothers and fathers. Level of marital distress was assessed by the Scale of Marriage Problems (Swenson and Fiore 1975), and a median split was used to divide the sample into distressed and nondistressed subgroups. Observation of parental teaching strategies and the child's responsiveness was accomplished with an event-recording procedure (Sigel et al. 1977) that yielded interrater reliability coefficients exceeding .75 for each of the eight behaviors coded. ANOVAs produced significant Marital Problems × Parent interactions for seven of the eight behavioral indices. Nondistressed mothers and fathers did not differ on any of the behavioral indices. By contrast, distressed mothers and fathers differed in their teaching strategies, the mothers' strategies being more effective: more questions, feedback, and suggestions and fewer attempts to take over the child's problem-solving efforts.

Fathers in the distressed group "behave in a more intrusive manner with their school-aged children, doing tasks for them rather than allowing them to discover their own solutions and displaying fewer positive emotions in response to their children's learning attempts" (p. 295). Mothers in distressed marriages, by contrast,

responded with more effective teaching behaviors, inducing more responsive behavior from their children. Hence the hypothesis of compensatory maternal behaviors in a distressed marriage was supported. The psychological basis for such compensation, however, remained conjectural, with the strong likelihood that mothers were compensating for perceived less-than-satisfactory parenting by their husbands. Finally, Brody et al. (1986) offered the caveat that the outcomes could not be generalized to parents with more meager educational and economic resources than characterized the well-educated parents employed in their study.

In two additional studies (Pellegrini et al. 1985, 1986), the Sigel research paradigm was applied, but interpretation of the results leaned heavily on Vygotsky's (1978) theory of the zone of proximal development. Pellegrini et al. (1985) studied parents' book-reading behaviors with 4- and 5-year-old children. Families differed in whether their children were communicatively disabled. MANOVA was applied, with the parental interaction behavior as the dependent variable and age, CD vs. non-CD status, and parent (mother vs. father) as the independent variables. Only CD vs. non-CD status yielded a significant main effect. Parental behaviors were more directive and less demanding with CD children. Furthermore, stepwise regression analysis examined the link between the parental interaction variables and WPPSI verbal IQ. For non-CD children, high cognitive demand was significantly associated with higher IQ levels; for CD children, the strongest positive predictor of IQ was the less demanding strategy of verbal/emotional support.

In general, parents seemed to adjust the cognitive demands of their teaching strategies to the level of the children's communicative competences. In Vygotskyan terms, parents operated within the child's zone of proximal development. Other evidence indicated that parents engaged in scaffolding to enhance their children's cognitive–linguistic performances. Thus parents of non-CD children manifested more conversational turns in a presumed effort to elicit more language from their children. Similarly, more parental paraphrasing with non-CD children encouraged departures from the literal text, thereby fostering greater depth of interaction between parent and child. In sum, parental scaffolding of their children's task-oriented behavior activated the potential for children to advance toward more independent problem solving as outlined in Vygotsky's theory.

We turn, finally, to the second study (Pellegrini et al. 1986) influenced by Vygotsky's theory. The research paradigm was similar to studies previously described. Again, gender of parent, children's CD vs. non-CD status, and the tasks of book reading and paper folding constituted the independent variables, and the teaching strategies of the parents comprised the dependent variables. In addition, the extent of task engagement by the child was also examined. MANOVA was employed, and it yielded a significant main effect for the child's communicative status and for its interaction with the task variable. ANOVAs applied to the separate teaching variables indicated that (a) parents were more directive and less demanding with CD children than with non-CD children; (b) parents were more demanding, gave less emotional support, and asked fewer questions with the paper-folding task than with the book-reading task; and (c) communicative status and task variable interacted: A CD versus non-CD difference occurred only for the book-reading task,

with parents of CD children asking more questions and making lower cognitive demands.

The teaching strategy measures were factor analyzed, and the resulting four orthogonal factors became the predictor variables in a regression analysis with children's rated task engagement as the criterion variable. For the paper-folding task, parents of both CD and non-CD children used high-demand strategies to keep their children engaged. For the book-reading task, parents of CD and non-CD children differed, with the CD parents using less demanding strategies and the non-CD parents using more demanding ones.

Pellegrini et al. (1986) had shown how ultimate problem-solving outcomes are of less significance than the processes by which such outcomes are achieved. Adult guidance is the key, with non-CD children requiring considerably less of it to remain engaged with the task than was the case for CD children. Hence the children's competence levels alert the parents to how demanding their teaching strategies should be. Pellegrini et al. further recommended the exploration of the sequence of parental teaching strategies, as parents found it necessary on occasion to switch from more demanding to less demanding strategies when the child encountered difficulty (see Wertsch et al. 1980). In sum, the findings strongly support the Vygotsky model of parents teaching children through the zone of proximal development and the adjustment of parental teaching consistent with the competence level of their children.

15.1.5 Practice

An important feature of Sigel's research program was linking research to practice (Renninger 2007). As Sigel (2006) noted,

> efforts to apply research to practice require acknowledging the inherent tensions of trying to validate theory and research in practical settings. They require stretching and/or adapting the root metaphors in which we have been trained so that collaborations between researchers and practitioners are the basis of research and any application of research to practice. (p. 1022)

The research on representational competence and psychological distance has had widespread impact, notably for early childhood education (Hyson et al. 2006) and cognitive behavior therapy (Beck 1967).

15.2 Parental Influences, Migration, and Measurement

Laosa's empirical work and his position papers spanned the psychological development of children, particularly Hispanics. His methodological contributions included test theory, especially as it relates to the assessment of minority children, and a standardized measure of parental teaching strategies. The major foci of Laosa's

work to be considered here are parental influences on children's development, the consequences of migration for their adjustment and growth, and the measurement of their ability.

15.2.1 Parental Influences

Parental influence on children's intellectual development has been a topic of long-standing interest to developmental psychologists (e.g., Clarke-Stewart 1977). A particular concern in Laosa's work was Hispanic children, given the gap in their academic achievement. His early research concerned maternal teaching. Unlike much of the previous work in that area, Laosa made direct observations of the mothers teaching their children, instead of relying on mothers' self-reports about interactions with their children, and distinguished between two likely SES determinants of their teaching: education and occupation. In a study of Hispanic mother–child dyads (Laosa 1978), mother's education correlated positively with praising and asking questions during the teaching and correlated negatively with modeling (i.e., the mother working on the learning task herself while the child observes). However, mother's occupation did not correlate with any of the teaching variables, and neither did father's occupation. Laosa speculated that the education-linked differences in teaching strategies account for the relationship between mothers' education and their children's intellectual development found in other research (e.g., Bradley et al. 1977). Subsequently, Laosa (1980b) also suggested that the more highly educated mothers imitate how they were taught in school.

In a follow-up study of Hispanic and non-Hispanic White mother–child dyads (Laosa 1980b), the two groups differed on most of the teaching variables. Non-Hispanic White mothers praised and asked questions more, and Hispanic mothers modeled, gave visual cues, directed, and punished or restrained more. However, when mothers' education was statistically controlled, the differences between the groups disappeared; controlling for mothers' or fathers' occupation did not reduce the differences.

In a third study, with the Hispanic mother–child dyads (Laosa 1980a), mother's field independence, assessed by the Embedded Figure Test (Witkin et al. 1971) and WAIS Block Design (Wechsler 1955), correlated positively with mother's asking questions and praising, and correlated negatively with mother's modeling. The correlations were reduced, but their pattern was similar when mother's education was statistically controlled. Laosa suggested that asking questions and praising are self-discovery approaches to learning that reflect field independence, whereas modeling is a concrete approach that reflects field dependence; hence mothers were using strategies that foster their own cognitive style in their children. Mother's teaching strategies, in fact, correlated modestly but inconsistently with the children's field independence, as measured by the Children's Embedded Figures Test (Witkin et al. 1971), WISC Block Design (Wechsler 1949a), and Human Figure Drawing (Harris 1963), another measure of field independence. Most of the teaching strategies had

scattered correlations with the Children's Embedded Figures Test and Block Design: positive correlations with asking questions and praising (field-independent strategies) and negative correlations with modeling, punishing or restraining, and giving visual cues (field-dependent strategies).

In Laosa's later research, a recurring topic was the impact of parents' education on their children's intellectual development; this line of work was presumably motivated by the influence of education in his maternal-teaching studies. Laosa (1982b) viewed parental education as impacting the parent–child interaction and presented a conceptual model of this interaction as the mediator between parent education and the child's development. He reported further analyses of the samples of Hispanic and non-Hispanic White mother–child dyads.

In one analysis, non-Hispanic White mothers and fathers read to their children more than did Hispanic parents. When parents' education was statistically controlled, the group difference disappeared, but controlling for parents' occupation did not reduce it. In addition, non-Hispanic mothers had higher *realistic* educational aspirations for their children ("*realistically*, how much education do you think your child will receive?"); this difference also disappeared when mothers' education was controlled but not when their occupation was controlled.

In another analysis, mother's education correlated positively in both the Hispanic and non-Hispanic White groups with mother's reading to the child, but father's education was uncorrelated with father's reading to the child in either group. Parent's occupation did not correlate with reading in the two groups. In both groups, mother's education also correlated positively with mother's educational aspirations for the child, but mother's occupation was uncorrelated.

Also, in an analysis of the Hispanic group, mother's education correlated positively with the child's ability to read or write before kindergarten, though father's education was uncorrelated. Parent's occupation was also uncorrelated with literacy. In addition, parent's education correlated positively with their use of English with the child; parent's occupation also correlated positively but weakly with English use.

Laosa argued that the set of findings, in total, suggests that the lower educational level of Hispanic parents produced a discontinuity between their children's home and school environments that adversely affected academic achievement.

He explored the consequences of these parental influences on the test performance of 3-year-olds in two studies. In the first study (Laosa 1982a), which targeted non-Hispanic White children, a path analysis was employed to assess the relationships, direct and indirect, between a host of family influences (e.g., mother's education and occupation, mother's reading to the child, nonparents in the household reading to the child, mother's teaching strategies) and performance on the Preschool Inventory (Caldwell 1970), a test of verbal, quantitative, and perceptual-motor skills for kindergarten children. A Mother's Socioeducational Values factor (defined by mother's education and occupation and mother's reading to child) was the strongest determinant of test performance. Less powerful determinants included nonparents in the household (probably older siblings) reading to the child and mother's use of modeling in teaching. Laosa highlighted two important and unanticipated findings:

the apparent influence of siblings and the substantial and positive influence of modeling, contrary to the conventional wisdom that verbal teaching strategies, such as asking questions, are superior to nonverbal ones, such as modeling.

In the second study (Laosa 1984) of Hispanic and non-Hispanic White children, the groups differed in their means on three of the five scales of the McCarthy Scales of Children's Abilities (McCarthy 1972): Verbal, Quantitative, and Memory. When a Sib Structure/Size factor (later-born child, many siblings) was statistically controlled, the group differences were unaffected. But when either a Language factor (mother uses English with child, child uses English with mother) or an SES factor (parents' education, father's occupation, household income) was controlled, the differences were reduced; when both factors were controlled, the differences were eliminated. The findings led Laosa to conclude that these early ethnic-group differences in ability were explainable by differences in SES and English-language usage.

15.2.2 Migration

In a series of white papers, Laosa reviewed and synthesized the extant research literature on the consequences of migration for children's adjustment and development, particularly Hispanic children, and laid out the salient issues (Laosa 1990, 1997, 1999). One theme was the need for—and the absence of—a developmental perspective in studying migration: "what develops, and *when, how,* and *why* it develops" (Laosa 1999, p. 370). The pioneering nature of this effort is underscored by the observation almost two decades later that migration is neglected by developmental psychology (Suárez-Orozco and Carhill 2008; Suárez-Orozco et al. 2008).

In a 1990 paper, Laosa proposed a multivariate, conceptual model that described the determinants of the adaptation of Hispanic immigrant children to the new society. Key features of the model were the inclusion of variables antedating immigration (e.g., sending community), moderator variables (e.g., receiving community), and mediating variables (e.g., child's perceptions and expectations) between the stresses of immigration and the outcomes.

In a complex, longitudinal survey of Puerto Rican migrants in New Jersey schools, Laosa (2001) found that the majority of the student body were Hispanic in 46% of the schools and were native speakers of Spanish in 31%. Additionally, the majority of the student body was eligible for free lunch in 77% of the schools and was from families on public assistance in 46%. Laosa concluded that the migrants faced considerable segregation by ethnicity or race as well as considerable isolation by language in high-poverty schools, factors with adverse consequences for the students' social and academic development.

15.2.3 Measurement

The measurement and evaluation of children's ability and achievement, particularly the unbiased assessment of minority children, has long been beset by controversies (see Laosa 1977; Oakland and Laosa 1977). These controversies were sparked in the 1960s and 1970s by the Coleman report (Coleman et al. 1966), which suggested that average differences in the academic performance of Black and White students are more affected by their home background than by their schools' resources, and by Jensen's (1969) review of research bearing on genetic and environmental influences on intelligence. He concluded that genetics is a stronger influence, which many observers interpreted as suggesting that the well-established disparity between Black and White children in their average scores on intelligence tests is largely genetic in origin. The upshot was widespread concerns that these tests are biased and calls for banning their use in schools. These arguments were reignited by *The Bell Curve* (Herrnstein and Murray 1994), which concluded that intelligence is mainly heritable. As Laosa (1996) noted, "thus, like a refractory strain of retrovirus, the issues tend to remain latent and from time to time resurge brusquely onto the fore of public consciousness" (p. 155).

In a 1977 paper, Laosa summarized the earlier controversies and other criticisms of testing and discussed alternatives to current testing practices that had been developed in response. The alternatives included constructing "culture-fair" tests "whose content is equally 'fair' or 'unfair' to different cultural groups" (p. 14), translating tests from English, using norms for subgroups, adjusting scores for test takers with deprived backgrounds, devising tests for subgroups (e.g., the BITCH, a vocabulary test based on Black culture; Williams 1972), using criterion-based interpretations of scores (i.e., how well a student achieves a specific objective) instead of norm-based interpretations (i.e., how well he or she does on the test relative to others), employing tests of specific abilities rather than global measures like IQ, and making observations of actual behavior. Laosa cautioned that these alternatives may also be problematic and would need to be carefully evaluated.

In a companion piece, Laosa, joined by Thomas Oakland of the University of Texas, Austin (Oakland and Laosa 1977), provided a comprehensive account of standards for minority-group testing that had been formulated by professional organizations, the government, and the courts. They argued for the need to consider these standards in testing minority-group children.

Laosa (1982c), in a subsequent paper on measurement issues in the evaluation of educational programs, specifically Head Start, delineated the concept of population validity and its applicability to program evaluation. Population validity deals with the question, "Do the results yielded by a given assessment technique have the same meaning when administered to persons of different sociocultural backgrounds?" (p. 512). Laosa discussed threats to population validity: familiarity (performance is influenced by familiarity with the task), communication, role relations (performance is influenced by the test taker's relationship with the tester), and situational (e.g., physical setting, people involved).

In another paper, Laosa (1991) explicated the links between population validity, cultural diversity, and professional ethics. As an illustration, he described a study by Bradley et al. (1989) of children in three ethnic groups, Black, Hispanic, and non-Hispanic White, matched on their HOME inventory (Caldwell and Bradley 1985) scores, a measure of the home environment. The HOME inventory scores correlated appreciably with performance on the Bayley Scales of Infant Development (Bayley 1969) and the Stanford–Binet Intelligence Test (Terman and Merrill 1960) for the Black and non-Hispanic White children but not for the Hispanic children. Laosa suggested that this finding highlights the importance of evaluating test results separately for different ethnic groups.

Laosa pointed out that population validity is a scientific concern in basic research and an ethical issue in applied work, given the inability to predict the results in different populations from the one studied. He also noted that when population differences are observed, two questions need to be answered. One, relevant to applied work, is, Which populations react differently? The other question, pertinent to scientific research, but rarely asked, is, Why do they differ?

In his last paper on measurement issues, Laosa (1996), responding to *The Bell Curve* controversy, made several general points about test bias. One was that bias reflects the absence of population validity. He noted that this view accords with the Cole and Moss (1989) definition of bias: "Bias is differential validity of a given interpretation of a test score for any definable, relevant group of test takers" (p. 205).

Another point was that the definition of predictive bias in the then current third edition of the *Standards for Educational and Psychological Testing* (American Educational Research Association, American Psychological Association, & National Council on Measurement in Education 1985) is insufficient. According to the *Standards*, predictive bias is absent if "the predictive relationship of two groups being compared can be adequately described by a common algorithm (e.g., regression line)" (p. 12). Laosa took up the argument that intelligence tests cannot be considered to be unbiased simply because their predictions are equally accurate for different races or social classes, noting Campbell's rejoinder that the same result would occur if the tests simply measured opportunity to learn (D. T. Campbell, personal communication, May 18, 1995).

The last point was that the consequences of test use also need to be considered (Messick 1989). Laosa cited Linn's (1989) example that requiring minimum high school grades and admissions test scores for college athletes to play during their freshman year can affect what courses minority athletes take in high school, whether they will attend college if they cannot play in their freshman year, and, ultimately, their education and employment.

15.3 Cognitive, Personality, and Social Development of Infants and Young Children

Lewis studied infant's cognitive attention and language ability, infants' and young children's physiological responses during attention, and infants' social and emotional development. He also formulated theories of development as well as theories about the integration of children's various competencies.

15.3.1 Social Development

Social development was a major interest, in particular, the mother–child interaction and the role this interaction played in the child's development. This work on social development revolved around four themes: (a) the mother–child relationship, (b) the growth of the child's social knowledge, (c) social cognition or the nature of the social world, and (d) the social network of children.

For example, in a 1979 volume (Lewis and Rosenblum 1979), *The Child and Its Family,* Lewis challenged the idea that the child's mother was the only important figure in the infant's early life and showed that fathers and siblings, as well as grandparents and teachers, were also key influences. And in a 1975 volume, on peer friendship in the opening years of life (Lewis and Rosenblum 1975), Lewis disputed the Piagetian idea that children could not form and maintain friendships before the age of 4 years. The finding that infants are attracted to and enjoy the company of other infants and young children, and that they can learn from them through observation and imitation, helped open the field of infant daycare (Goldman and Lewis 1976; Lewis 1982b; Lewis and Schaeffer 1981; Lewis et al. 1975). Because the infant's ability to form and maintain friends is important for the daycare context, where groups of infants are required to play together, this work also showed that the learning experience of young children and infants involved both direct and indirect interactions, such as modeling and imitation with their social world of peers, siblings, and teachers (Feiring and Lewis 1978; Lewis and Feiring 1982). This information also had an important consequence on hospital care; until this time, infants were kept far apart from each other in the belief that they could not appreciate or profit from the company of other children.

Another major theme of the research on social development involved infants' social knowledge. In a series of papers, Lewis was able to show that infants could discriminate among human faces (Lewis 1969); that they were learning about their gender (Lewis and Brooks 1975); that they were showing sex-role-appropriate behaviors (Feiring and Lewis 1979; Goldberg and Lewis 1969; Lewis 1975a); that they were learning about how people look, for example, showing surprise at the appearance of a midget—a child's height but an adult's face (Brooks and Lewis 1976); and that they were detecting the correspondence between particular faces and voices (McGurk and Lewis 1974). All of these results indicated that in the first

2 years, children were learning a great deal about their social worlds (Brooks-Gunn and Lewis 1981; Lewis 1981b; Lewis et al. 1971).

The most important aspect of this work on social development was the child's development of a sense of itself, something now called consciousness, which occurs in the second and third years of life. In the Lewis and Brooks-Gunn (1979a) book on self-recognition, *Social Cognition and the Acquisition of Self,* Lewis described his mirror self-recognition test, a technique that has now been used across the world. Results with this test revealed that between 15 and 24 months of age, typically developing children come to recognize themselves in mirrors. He subsequently showed that this ability, the development of the idea of "me," along with other cognitive abilities gives rise to the complex emotions of empathy, embarrassment, and envy as well as the later self-conscious emotions of shame, guilt, and pride (Lewis and Brooks 1975; Lewis and Brooks-Gunn 1981b; Lewis and Michalson 1982b; Lewis and Rosenblum 1974b).

These ideas, an outgrowth of the work on self-recognition, led to Lewis's interest in emotional development. They also resulted in a reinterpretation of the child's need for others. While children's attachment to their mothers was thought to be the most important relationship for the children, satisfying all of their needs, it became clear that others played an important role in children's social and emotional lives. His empirical work on fathers (Lewis and Weinraub 1974) and peers (Lewis et al. 1975) led to the formulation of a social network theory (Feiring and Lewis 1978; Lewis 1980; Lewis and Ban 1977; Lewis and Feiring 1979; Weinraub et al. 1977).

15.3.2 Emotional Development

Lewis's interest in social development and in consciousness led quite naturally to his research on emotional development, as already noted (Lewis 1973, 1977b, 1980; Lewis and Brooks 1974; Lewis et al. 1978; Lewis and Michalson 1982a, b; Lewis and Rosenblum 1978a, b). Two volumes framed this work on the development of emotional life (Lewis and Rosenblum 1974b, 1978b) and were the first published studies of emotional development. These early efforts were focused on the emotions of infants in the first year of life, including fear, anger, sadness, joy, and interest. To study emotional life, Lewis created experimental paradigms and devised a measurement system. So, for example, paradigms were developed for peer play (Lewis et al. 1975), social referencing (Feinman and Lewis 1983; Lewis and Feiring 1981), stranger approach (Lewis and Brooks-Gunn 1979a), mirror recognition (Lewis and Brooks-Gunn 1979a), and contingent learning (Freedle and Lewis 1970; Lewis and Starr 1979). A measurement system was created for observing infants' and young children's emotional behavior in a daycare situation that provided scales of emotional development (Lewis and Michalson 1983). These scales have been used by both American and Italian researchers interested in the effects of daycare on emotional life (Goldman and Lewis 1976).

15.3.3 Cognitive Development

Lewis's interests in development also extended to the study of infants' and children's cognitive development, including attentional processes, intelligence, and language development (Dodd and Lewis 1969; Freedle and Lewis 1977; Hale and Lewis 1979; Lewis 1971b, 1973, 1975b, 1976a, b, 1977a, 1978a, 1981a, 1982a; Lewis and Baldini 1979; Lewis and Baumel 1970; Lewis and Cherry 1977; Lewis and Freedle 1977; Lewis and Rosenblum 1977; Lewis et al. 1969a, 1971; McGurk and Lewis 1974).

Lewis first demonstrated that the Bayley Scales of Infant Development (Bayley 1969), which were—and still are—the most widely used test of infant intelligence, had no predictive ability up to 18 months of age (Lewis and McGurk 1973). In an effort to find an alternative, Lewis turned to research on infants' attentional ability, which he had begun at the Fels Research Institute, and developed it further at ETS. This work used a habituation–dishabituation paradigm where the infant was presented with the same visual stimulus repeatedly and then, after some time, presented with a variation of that stimulus. Infants show boredom to the repeated event, or habituation, and when the new event is presented, the infants show recovery of their interest, or dishabituation (Kagan and Lewis 1965; Lewis et al. 1967a, b). Infants' interest was measured both by observing their looking behavior and by assessing changes in their heart rate (Lewis 1974; Lewis et al. 1966a, b; Lewis and Spaulding 1967). He discovered that the infants' rate of habituation and degree of dishabituation were both related to their subsequent cognitive competence, in particular to their IQ. In fact, this test was more accurate than the Bayley in predicting subsequent IQ (Lewis and Brooks-Gunn 1981a, c; Lewis et al. 1969; Lewis and McGurk 1973).

This research on attentional processes convinced Lewis of the usefulness of physiological measures, such as heart rate changes, in augmenting behavior observation, work which he also began at the Fels Research Institute and continued and expanded at ETS (Lewis 1971a, b, 1974; Lewis et al. 1969, 1970, 1978; Lewis and Taft 1982; Lewis and Wilson 1970; Sontag et al. 1969; Steele and Lewis 1968).

15.3.4 Atypical Development

Lewis's research on normal development, especially on attentional processes as a marker of central nervous system functioning, led to an interest in atypical developmental processes and a program of research on children with disabilities (Brinker and Lewis 1982a, b; Brooks-Gunn and Lewis 1981, 1982a, b, c; Fox and Lewis 1982a, b; Lewis 1971c; Lewis and Fox 1980; Lewis and Rosenblum 1981; Lewis and Taft 1982; Lewis and Wehren 1982; Lewis and Zarin-Ackerman 1977; Thurman and Lewis 1979; Zarin-Ackerman et al. 1977). Perhaps of most importance was the development of an intervention strategy based on Lewis's work with typically

developing children, the Learning to Learn Curriculum. Infants with disabilities were given home- and clinic-based interventions where their simple motor responses resulted in complex outcomes and where they had to learn to produce these outcomes, which served as operants—in effect, an applied-behavior-analysis intervention strategy (Brinker and Lewis 1982a, b; Lewis 1978a, b; Thurman and Lewis 1979).

15.3.5 Theories

Lewis formulated several influential theories about infant development. These included (a) a reconsideration of attachment theory (Weinraub and Lewis 1977) and (b) the infant as part of a social network (Weinraub et al. 1977). He also began work on a theory of emotional development (Lewis 1971b; Lewis and Michalson 1983).

15.3.6 The Origin of Behavior Series

Lewis and Leonard Rosenblum of SUNY Downstate Medical Center organized yearly conferences on important topics in child development for research scientists in both child and animal (primate) development to bring together biological, cultural, and educational points of view. These conferences resulted in a book series, *The Origins of Behavior* (later titled *Genesis of Behavior*), under their editorship, with seven highly cited volumes (Lewis and Rosenblum 1974a, b, 1975, 1977, 1978a, 1979, 1981). The initial volume, *The Effect of the Infant on the Caregiver* (Lewis and Rosenblum 1974a), was so influential that the term *caregiver* became the preferred term, replacing the old term *caretaker.* The book became the major reference on the interactive nature of social development—that the social development of the child involves an interaction between the mother's effect on the infant and the effect of the infant on the mother. It was translated into several languages, and 15 years after publication, a meeting sponsored by the National Institutes of Health reviewed the effects of this volume on the field.

15.4 Cognitive, Personality, and Social Development From Infancy to Adolescence

Brooks-Gunn's work encompassed research on the cognitive, personality, and social development of infants, toddlers, and adolescents, primarily within the framework of social-cognitive theory. Major foci were the acquisition of social knowledge in young children, reproductive processes in adolescence, and perinatal influences on

children's development. These issues were attacked in laboratory experiments, other cross-sectional and longitudinal studies, and experimental interventions. (A fuller account appears in Brooks-Gunn 2013.)

15.4.1 Social Knowledge in Infants and Toddlers

In collaboration with Lewis, Brooks-Gunn carried out a series of studies on the development of early knowledge about the self and others in infancy and toddlerhood. They investigated how and when young children began to use social categories, such as gender, age, and relationship, to organize their world and to guide interactions (Brooks and Lewis 1976; Brooks-Gunn and Lewis 1979a, b, 1981) as well as the development of self-recognition as a specific aspect of social cognition (Lewis and Brooks-Gunn 1979b, c; Lewis et al. 1985). This developmental work was embedded in genetic epistemology theory as well as social-cognitive theory, with a strong focus on the idea that the self (or person) only develops in relation to others and that the self continues to evolve over time, as does the relation to others.

The studies demonstrated that social knowledge develops very early. Infants shown pictures of their parents, strange adults, and 5-year olds and asked, Who is that? were able to label their parents' pictures as mommy and poppy, labeling their fathers' pictures more accurately and earlier than their mothers' (Brooks-Gunn and Lewis 1979b). Shown pictures of their parents and strange adults, infants smiled more often and looked longer at their parents' pictures (Brooks-Gunn and Lewis 1981). And when infants were approached by strangers—5-year-old boys and girls, adult men and women, and a midget woman—the children discriminated among them on the basis of age and height, smiling and moving toward the children but frowning and moving away from the adults and, compared to the other adults, watching the midget more intently and averting their gaze less (Brooks and Lewis 1976).

15.4.2 Reproductive Events

15.4.2.1 Menstruation and Menarche

Brooks-Gunn's interest in the emergence of social cognition broadened to its role in the development of perceptions about reproductive events, at first menstruation and menarche. Her focus was on how social cognitions about menstruation and menarche emerge in adolescence and how males' and females' cognitions differ. Brooks-Gunn and Diane Ruble, then at Princeton University, began a research program on the salience and meaning of menarche and menstruation, especially in terms of definition of self and other in the context of these universal reproductive events

(Brooks-Gunn 1984, 1987; Brooks-Gunn and Ruble 1982a, b, 1983; Ruble and Brooks-Gunn 1979b). They found that menstruation was perceived as more physiologically and psychologically debilitating and more bothersome by men than by women (Ruble et al. 1982). In addition, their research debunked a number of myths about reproductive changes (Ruble and Brooks-Gunn 1979a), including the one that menarche is a normative crisis experienced very negatively by all girls. In fact, most girls reported mixed emotional reactions to menarche that were quite moderate. These reactions depended on the context the girls experienced: Those who were unprepared for menarche or reached it early reported more negative reactions as well as more symptoms (Ruble and Brooks-Gunn 1982).

15.4.2.2 Pubertal Processes

Brooks-Gunn's research further broadened to include pubertal processes. With Michelle Warren, a reproductive endocrinologist at Columbia University, she initiated a research program on pubertal processes and the transition from childhood to early adolescence. Brooks-Gunn and Warren conducted longitudinal studies of girls to chart their emotional experiences associated with pubertal changes and the socialization practices of families. The work included measurement of hormones to better understand pubertal changes and possible emotional reactions. The investigations followed girls who were likely to have delayed puberty because of exercise and food restriction (dancers training in national ballet companies as well as elite swimmers and gymnasts) and girls attending private schools—many of the girls were followed from middle school through college (Brooks-Gunn and Warren 1985, 1988a, b; Warren et al. 1986, 1991).

The private-school girls commonly compared their pubertal development and had no difficulty categorizing their classmates' development (Brooks-Gunn et al. 1986). Relatedly, the onset of breast development for these girls correlated positively with scores on measures of peer relationships, adjustment, and body image, but pubic hair was uncorrelated, suggesting that breast growth may be a visible sign of adulthood, conferring enhanced status (Brooks-Gunn and Warren 1988b).

The context in which the girls were situated influenced their reactions. In a context where delayed onset of puberty is valued (the dance world—most professional ballerinas are late maturers), dancers with delayed puberty had higher scores (relative to on-time dancers) on a body-image measure (Brooks-Gunn, Attie, Burrow, Rosso, & Warren, Brooks-Gunn et al. 1989; Brooks-Gunn and Warren 1985). (They also had lower scores on measures of psychopathology and bulimia; Brooks-Gunn and Warren 1985.) In contrast, in contexts where delayed onset is not valued (swimmers, private-school students/nonathletes), delayed and on-time girls did not differ in their body images (Brooks-Gunn, Attie et al., Brooks-Gunn et al. 1989; Brooks-Gunn and Warren 1985).

Two publications in this program, in particular, were very widely cited, according to the Social Science Citation Index: Attie and Brooks-Gunn (1989), on eating

problems, and Brooks-Gunn et al. (1987), on measuring pubertal status, with 389 and 323 citations through 2015, respectively.

15.4.2.3 Adolescent Parenthood

Given Brooks-Gunn's research interest in menarche and other pubertal processes, it is not surprising that she moved on to research on pregnancy and parenthood, events that presage changes in self-definition as well as social comparisons with others. Brooks-Gunn joined Frank Furstenberg, a family sociologist at the University of Pennsylvania, in a 17-year follow-up of a group of teenage mothers who gave birth in Baltimore in the early 1970s (Furstenberg et al. 1987a, b, 1990). They charted the trajectories of these mothers as well as their children, who were about the age that their mothers had been when they gave birth to them. The interest was in both how well the mothers were doing and how the mothers' life course had influenced their children.

In brief, the teenage mothers differed widely in their life chances: About one third were on welfare and three quarters had jobs, usually full-time ones. Characteristics of the mothers' family of origin and of their own families (e.g., higher levels of education) and their attendance at a school for pregnant teenagers predicted the mothers' economic success.

The outcomes for their teenage children were "strikingly poor" (Brooks-Gunn 1996, p. 168). About one third were living with their biological father or stepfather. Half had repeated at least one grade in school, and most were sexually active. Maternal characteristics were linked to the children's behavior. Children of mothers who had not graduated from high school were 2.4 times as likely as other children, and children of unmarried mothers were 2.2 times as likely, to have repeated a grade. And children of unmarried mothers were 2.4 times as likely to have been stopped by the police, according to their mothers.

The Furstenberg et al. (1987b) monograph chronicling this study, *Adolescent Mothers in Later Life,* won the William J. Goode Book Award from the American Sociological Association's Sociology of the Family Section and is considered one of the classic longitudinal studies in developmental psychology.

Brooks-Gunn and Lindsay Chase-Lansdale, then at George Washington University, also began a study of low-income, Black multigenerational families (grandmother/grandmother figure–young mother–toddler) in Baltimore to investigate family relationships, via home visits (Chase-Lansdale et al. 1994). One issue was the parenting by the grandmother and mother, as observed separately in videotaped interactions of them aiding the child in working on a puzzle. The quality of parenting depended on whether they resided together and on the mother's age. Mothers' parenting was lower in quality when they lived with grandmothers. (Residing with the grandmothers and sharing child caring may be stressful for the mothers, interfering with their parenting.) Grandmothers' parenting was *higher* in quality when they lived with younger mothers than when they lived apart, but it was *lower* in quality when they lived with older mothers rather than apart. (Coresiding grandmothers may be more willing to help

younger mothers, whom they view as needing assistance in parenting, than older mothers, whom they see as capable of parenting on their own.)

15.4.3 Perinatal Influences

Another line of research expanded beyond teenage parents to look at perinatal conditions, such as low birth weight and pregnancy behavior (e.g., smoking, no prenatal care), that influence parenting and children's development. Poor families and mothers with low education were often the focus of this research, given the differential rates of both early parenthood and adverse perinatal conditions as a function of social class.

In a joint venture between ETS, St. Luke's–Roosevelt Hospital, and Columbia University's College of Physicians and Surgeons, Brooks-Gunn studied low-birth-weight children and their parents, many from disadvantaged families because of the greater incidence of low-birth-weight children in these families. The work led to her thinking about how to ameliorate cognitive, emotional, and academic problems in these vulnerable children (Brooks-Gunn and Hearn 1982).

Brooks-Gunn joined Marie McCormick, a pediatrician then at the University of Pennsylvania, in a 9-year follow-up of low-birth-weight infants from previous multisite studies (Klebanov et al. 1994; McCormick et al. 1992). The focus was on very low birth weight infants, for more of them were surviving because of advances in neonatal intensive care.

At age 9, the low-birth-weight children did not differ from normal-birth-weight children on most aspects of classroom behavior, as reported by their teachers, but they had lower attention/ language skills and scholastic competence and higher daydreaming and hyperactivity; these differences were most pronounced for extremely low birth weight children. This pattern of differences resembles attention deficit disorder (Klebanov et al. 1994). The low-birth-weight children also had lower mean IQs and, at home, more behavioral problems, as reported by their mothers. The adverse health status of these children underscores the importance of efforts to reduce the incidence of premature births (McCormick et al. 1992).

15.4.4 Interventions With Vulnerable Children

15.4.4.1 Low-Birth-Weight Children

Brooks-Gunn and McCormick also collaborated on two other research programs involving interventions with biologically vulnerable children, the majority of whom were poor. One program focused on reducing the incidence of low-birth-weight deliveries by providing pregnant women with child-rearing and health information.

This program used a public health outreach model to locate pregnant women who were not enrolled in prenatal care; the intervention was located at Harlem Hospital. This effort was a logical extension of Brooks-Gunn's work on adolescent mothers in Baltimore (Brooks-Gunn et al. 1989; McCormick et al. 1987, 1989a, b).

The women in the program were very disadvantaged: One fifth were adolescents, three quarters were single, and half had not graduated from high school. The birth weight of their infants was unrelated to traditional risk factors: mother's demographic (e.g., education) and psychosocial characteristics (e.g., social support). This outcome suggests that low birth weight *in poor populations* is largely due to poverty per se. Birth weight *was* associated with the adequacy of prenatal care (Brooks-Gunn et al. 1988; McCormick et al. 1987).

The outreach program was extensive—four local people searching for eligible women over the course of a year, each making roughly 20 to 25 contacts daily—but recruited only 52 additional women, at a cost of about $850 each. The labor-intensive and expensive nature of this outreach effort indicates that more cost-effective alternatives are needed (Brooks-Gunn et al. 1988, 1989).

The other program involved the design and implementation of an early intervention for premature, low-birth-weight infants: enrollment in a child development education center and family support sessions. This program was initiated in eight sites and included almost 1000 children and their families; randomization was used to construct treatment and control groups. These children were followed through their 18th year of life, with the intervention from birth to 3 years of age being evaluated by Brooks-Gunn (Infant Health and Development Program 1990). The 3-year-olds in the intervention group had higher mean IQs and fewer maternally reported behavior problems, suggesting that early intervention may decrease low-birth-weight infants' risk of later developmental disability.

15.4.4.2 Head Start

Brooks-Gunn also carried out a notable evaluation of Head Start, based on data from an earlier longitudinal study conducted at ETS in the 1970s. The ETS–Head Start Longitudinal Study, directed by Shipman (1972, 1973), had canvassed poor school districts in three communities in an effort to identify and recruit for the study all children who were 3 ½ to 4 ½ years old, the Head Start population. The children were then assessed and information about their families was obtained. They were reassessed annually for the next 3 years. After the initial assessment, some children had entered Head Start, some had gone to other preschool programs, and some had not enrolled in any program. Clearly families chose whether to enroll their children in Head Start, some other program, or none at all (by processes that are difficult if not impossible to measure). But, by having the children's assessments and familial and demographic measures at age 3, it was possible to document and control statistically for initial differences among children and families in the three groups. Children's gains in ability in these groups could then be compared.

In several studies of two communities (Lee et al. 1988, 1990; Schnur et al. 1992), Brooks-Gunn and her collaborators investigated differences in the children's gains in the Head Start and other groups as well as preexisting group differences in the children's demographic and cognitive characteristics. Black children enrolled in Head Start made greater gains on a variety of cognitive tests than their Black peers in the other groups by the end of the program (Lee et al. 1988) and diminished gains after 2 years (Lee et al. 1990). (The gains for the small samples of White children did not differ between the Head Start and other groups in the initial study; these children were not included in the follow-up study.) These findings imply that Head Start may have some efficacy in improving participants' intellectual status. The Head Start children were the most disadvantaged (Schnur et al. 1992), seemingly allaying concerns that Head Start does not take the neediest children (Datta 1979).

15.5 Conclusion

As this review documents, ETS was a major center for basic and applied research in developmental psychology for decades. The number and quality of investigators (and their prodigious output) made for a developmental psychology program that rivaled the best in the academic community.

The research was wide ranging and influential, spanning the cognitive, personality, and social development of infants, children, and adolescents, with an emphasis on minority, working-class, and disabled individuals; addressing key theoretical, substantive, and methodological issues; using research methods that ran the gamut: laboratory and field experiments, correlational studies, surveys, and structured observations; and impacting theory, research, and practice across developmental psychology.

In common with ETS's research in cognitive, personality, and social psychology (Stricker, Chap. 13, and Kogan, Chap. 14, this volume), this achievement was probably attributable to the confluence of ample institutional and financial support, doubtless due to the vision of Chauncey, who saw the value of a broad program of psychological research.

Acknowledgments Thanks are due to Nick Telepak for retrieving publications and to Isaac Bejar, Randy Bennett, and Ann Renninger for reviewing a draft of this chapter.

References

Ajzen, I., & Fishbein, M. (1977). Attitude–behavior relations: A theoretical analysis and review of empirical research. *Psychological Bulletin, 84*, 888–918. https://doi.org/10.1037/0033-2909.84.5.888

American Educational Research Association, American Psychological Association, & National Council on Measurement in Education. (1985). *Standards for educational and psychological testing*. Washington, DC: American Psychological Association.

Attie, S., & Brooks-Gunn, J. (1989). Development of eating problems in adolescent girls: A longitudinal study. *Developmental Psychology, 25*, 70–79. https://doi.org/10.1037/0012-1649.25.1.70

Bayley, N. (1969). *Manual for the Bayley Scales of Infant Development*. New York: Psychological Corporation.

Beck, A. T. (1967). *Depression: Clinical, experimental, and theoretical aspects*. New York: Harper & Row.

Bradley, R. H., Caldwell, B. M., & Elardo, R. (1977). Home environment, social status, and mental test performance. *Journal of Educational Psychology, 69*, 697–701. https://doi.org/10.1037/0022-0663.69.6.697

Bradley, R. H., Caldwell, B. M., Rock, S. L., Ramey, C. T., Barnard, K. E., Gray, C., et al. (1989). Home environment and cognitive development in the first 3 years of life: A collaborative study involving six sites and three ethnic groups in North America. *Developmental Psychology, 25*, 217–235. https://doi.org/10.1037/0012-1649.25.2.217

Brinker, R. P., & Lewis, M. (1982a). Discovering the competent handicapped infant: A process approach to assessment and intervention. *Topics in Early Childhood Special Education, 2*(2), 1–16. https://doi.org/10.1177/027112148200200205

Brinker, R. P., & Lewis, M. (1982b). Making the world work with microcomputers: A learning prosthesis for handicapped infants. *Exceptional Children, 49*, 163–170. https://doi.org/10.1177/001440298204900210

Brody, G. H., Pellegrini, A. D., & Sigel, I. E. (1986). Marital quality and mother–child and father–child interactions with school-aged children. *Developmental Psychology, 22*, 291–296. https://doi.org/10.1037/0012-1649.22.3.291

Brooks, J., & Lewis, M. (1976). Infants' responses to strangers: Midget, adult, and child. *Child Development, 47*, 323–332. https://doi.org/10.2307/1128785

Brooks-Gunn, J. (1984). The psychological significance of different pubertal events in young girls. *Journal of Early Adolescence, 4*, 315–327. https://doi.org/10.1177/0272431684044003

Brooks-Gunn, J. (1987). Pubertal processes and girls' psychological adaptation. In R. Lerner & T. T. Foch (Eds.), *Biological–psychosocial interactions in early adolescence* (pp. 123–153). Hillsdale: Erlbaum.

Brooks-Gunn, J. (1996). Unexpected opportunities: Confessions of an eclectic developmentalist. In M. R. Merrens & G. C. Brannigan (Eds.), *The developmental psychologists—Research adventures across the life span* (pp. 153–171). New York: McGraw-Hill.

Brooks-Gunn, J. (2013). Person, time, and place—The life course of a developmental psychologist. In R. M. Lerner, A. C. Petersen, R. K. Silbereisen, & J. Brooks-Gunn (Eds.), *The developmental science of adolescence—History through autobiography* (pp. 32–44). New York: Psychology Press.

Brooks-Gunn, J., Attie, I., Burrow, C., Rosso, J. T., & Warren, M. P. (1989). The impact of puberty on body and eating concerns in athletic and nonathletics contexts. *Journal of Early Adolescence, 9,* 269–290. https://doi.org/10.1177/0272431689093006

Brooks-Gunn, J., & Hearn, R. P. (1982). Early intervention and developmental dysfunction: Implications for pediatrics. *Advances in Pediatrics, 29*, 497–527.

Brooks-Gunn, J., & Lewis, M. (1979a). The effects of age and sex on infants' playroom behavior. *Journal of Genetic Psychology, 134*, 99–105. https://doi.org/10.1080/00221325.1979.10533403

Brooks-Gunn, J., & Lewis, M. (1979b). "Why mama and papa?" The development of social labels. *Child Development, 50*, 1203–1206. https://doi.org/10.2307/1129349

Brooks-Gunn, J., & Lewis, M. (1981). Infant social perceptions: Responses to pictures of parents and strangers. *Developmental Psychology, 17*, 647–649. https://doi.org/10.1037/0012-1649.17.5.647

Brooks-Gunn, J., & Lewis, M. (1982a). Affective exchanges between normal and handicapped infants and their mothers. In T. Field & A. Fogel (Eds.), *Emotion and early interaction* (pp. 161–188). Hillsdale: Erlbaum.

Brooks-Gunn, J., & Lewis, M. (1982b). Development of play behavior in handicapped and normal infants. *Topics in Early Childhood Special Education, 2*(3), 14–27. https://doi.org/10.1177/027112148200200306

Brooks-Gunn, J., & Lewis, M. (1982c). Temperament and affective interaction in handicapped infants. *Journal of Early Intervention, 5*, 31–41. https://doi.org/10.1177/105381518200500105

Brooks-Gunn, J., McCormick, M. C., Gunn, R. W., Shorter, T., Wallace, C. Y., & Heagarty, M. C. (1989). Outreach as case finding—The process of locating low-income pregnant women. *Medical Care, 27,* 95–102. https://doi.org/10.1097/00005650-198902000-00001

Brooks-Gunn, J., & Ruble, D. N. (1982a). Developmental processes in the experience of menarche. In A. Baum & J. E. Singer (Eds.), *Hanbook of psychology and health: Vol. 2. Issues in child health and adolescent health* (pp. 117–147). Hillsdale: Erlbaum.

Brooks-Gunn, J., & Ruble, D. N. (1982b). The development of menstrual beliefs and behavior during early adolescence. *Child Development, 53*, 1567–1577. https://doi.org/10.2307/1130085

Brooks-Gunn, J., & Ruble, D. N. (1983). The experience of menarche from a developmental perspective. In J. Brooks-Gunn & A. C. Petersen (Eds.), *Girls at puberty—Biological and psychosocial perspectives* (pp. 155–177). New York: Plenum Press. https://doi.org/10.1007/978-1-4899-0354-9_8

Brooks-Gunn, J., & Warren, M. P. (1985). The effects of delayed menarche in different contexts: Dance and nondance students. *Journal of Youth and Adolescence, 141*, 285–300. https://doi.org/10.1007/BF02089235

Brooks-Gunn, J., & Warren, M. P. (1988a). Mother–daughter differences in menarcheal age in adolescent girls attending national dance company schools and non-dancers. *Annals of Human Biology, 15*, 35–43. https://doi.org/10.1080/03014468800009441

Brooks-Gunn, J., & Warren, M. P. (1988b). The psychological significance of secondary sexual characteristics in nine- to eleven-year-old girls. *Child Development, 59*, 1061–1069. https://doi.org/10.2307/1130272

Brooks-Gunn, J., Warren, M. P., Samelson, M., & Fox, R. (1986). Physical similarity to and disclosure of menarcheal status to friends: Effects of grade and pubertal status. *Journal of Early Adolescence, 6*, 3–34. https://doi.org/10.1177/0272431686061001

Brooks-Gunn, J., Warren, M. P., Rosso, J., & Gargiuolo, J. (1987). Validity of self-report measures of girls' pubertal status. *Child Development, 58*, 829–841. https://doi.org/10.2307/1130220

Brooks-Gunn, J., McCormick, M. C., & Heagarty, M. C. (1988). Preventing infant mortality and morbidity: Developmental perspectives. *American Journal of Orthopsychiatry, 58*, 288–296. https://doi.org/10.1111/j.1939-0025.1988.tb01590.x

Caldwell, B. M. (1970). *Handbook, Preschool Inventory* (Rev. ed.). Princeton: Educational Testing Service.

Caldwell, B., & Bradley, R. (1985). *Home observation for measurement of the environment*. Homewood: Dorsey.

Chase-Lansdale, P. L., Brooks-Gunn, J., & Zamsky, E. S. (1994). Young African-American multigenerational families in poverty: Quality of mothering and grandmothering. *Child Development, 65*, 373–393. https://doi.org/10.2307/1131390

Clarke-Stewart, A. (1977). *Child care in the family—A review of research and some propositions for policy*. New York: Academic Press.

Cole, N. S., & Moss, P. A. (1989). Bias in test use. In R. L. Linn (Ed.), *Educational measurement* (3rd ed., pp. 201–219). New York: Macmillan.

Coleman, J. S., Campbell, E. Q., Hobson, C. J., McPartland, J., Mood, A. M., Weinfeld, F. D., & York, R. L. (1966). *Equality of educational opportunity*. Washington, DC: U.S. Department of Health, Education, and Welfare, Office of Education.

Copple, C., Sigel, I. E., & Saunders, R. A. (1984). *Educating the young thinker—Classroom strategies for cognitive growth*. Hillsdale: Erlbaum.

Datta, L.-E. (1979). Another spring and other hopes: Some findings from national evaluations of Project Head Start. In E. Zigler & J. Valentine (Eds.), *Project Head Start—A legacy of the war on poverty* (pp. 405–432). New York: Free Press.

Dodd, C., & Lewis, M. (1969). The magnitude of the orienting response in children as a function of changes in color and contour. *Journal of Experimental Child Psychology, 8*, 296–305. https://doi.org/10.1016/0022-0965(69)90104-0

Emmerich, W. (1968). Personality development and the concept of structure. *Child Development, 39*, 671–690. https://doi.org/10.2307/1126978

Emmerich, W. (1982). Development of sex-differentiated preferences during late childhood and adolescence. *Developmental Psychology, 18*, 406–417. https://doi.org/10.1037/0012-1649.18.3.406

Feinman, S., & Lewis, M. (1983). Social referencing at ten months: A second-order effect on infants' responses to strangers. *Child Development, 50*, 848–853. https://doi.org/10.2307/1129892

Feiring, C., & Lewis, M. (1978). The child as a member of the family system. *Behavioral Science, 23*, 225–233. https://doi.org/10.1002/bs.3830230311

Feiring, C., & Lewis, M. (1979). Sex and age differences in young children's reactions to frustration: A further look at the Goldberg and Lewis subjects. *Child Development, 50*, 848–853. https://doi.org/10.2307/1128953

Fox, N., & Lewis, M. (1982a). Motor asymmetries in preterm infants: Effects of prematurity and illness. *Developmental Psychobiology, 15*, 19–23. https://doi.org/10.1002/dev.420150105

Fox, N., & Lewis, M. (1982b). Prematurity, illness, and experience as factors in preterm development. *Journal of Early Intervention, 6*, 60–71. https://doi.org/10.1177/105381518200600108

Freedle, R., & Lewis, M. (1970). On relating an infant's observation time of visual stimuli with choice-theory analysis. *Developmental Psychology, 2*, 129–133. https://doi.org/10.1037/h0028601

Freedle, R., & Lewis, M. (1977). Prelinguistic conversations. In M. Lewis & L. A. Rosenblum (Eds.), *The origins of behavior: Vol. 5. Interaction, conversation, and the development of language* (pp. 157–185). New York: Wiley.

Furstenberg Jr., F. F., Brooks-Gunn, J., & Morgan, S. P. (1987a). Adolescent mothers and their children in later life. *Family Planning Perspectives, 19*, 142–151. https://doi.org/10.2307/2135159

Furstenberg Jr., F. F., Brooks-Gunn, J., & Morgan, S. P. (1987b). *Adolescent mothers in later life*. New York: Cambridge University Press.

Furstenberg Jr., F. F., Levine, J. A., & Brooks-Gunn, J. (1990). The children of teenage mothers: Patterns of early childbearing in two generations. *Family Planning Perspectives, 22*, 54–61. https://doi.org/10.2307/2135509

Goldberg, S., & Lewis, M. (1969). Play behavior in the year-old infant: Early sex differences. *Child Development, 40*, 21–31. https://doi.org/10.2307/1127152

Goldman, K. S., & Lewis, M. (1976). *Child care and public policy: A case study*. Princeton: Educational Testing Service.

Hale, G. A., & Alderman, L. B. (1978). Children's selective attention and variation in amount of stimulus exposure. *Journal of Experimental Child Psychology, 25*, 320–327. https://doi.org/10.1016/0022-0965(78)90011-5

Hale, G. A., & Lewis, M. (Eds.). (1979). *Attention and cognitive development*. New York: Plenum Press. https://doi.org/10.1007/978-1-4613-2985-5

Harris, D. B. (1963). *Children's drawings as measures of intellectual maturity*. New York: Harcourt, Brace & World.

Herrnstein, R. J., & Murray, C. (1994). *The bell curve—Intelligence and class structure in American life*. New York: Free Press.

Hyson, M., Copple, C., & Jones, J. (2006). Early childhood development and education. In W. Damon & R. M. Lerner (Series Eds.), K. A. Renninger & I. E. Sigel (Vol. Eds.), *Handbook of child psychology: Vol. 4. Child psychology in practice* (6th ed., pp. 3–47). New York: Wiley.

Infant Health and Development Program. (1990). Enhancing the outcomes of low-birth-weight, premature infants: A multisite, randomized trial. *Journal of the American Medical Association, 263*, 3035–3042. https://doi.org/10.1001/jama.1990.03440220059030

Jensen, A. R. (1969). How much can we boost IQ and scholastic achievement? *Harvard Educational Review, 39*, 1–123. https://doi.org/10.17763/haer.39.1.l3u15956627424k7

Kagan, J., & Lewis, M. (1965). Studies of attention in the human infant. *Merrill-Palmer Quarterly, 11*, 95–127.

Klebanov, P. K., Brooks-Gunn, J., & McCormick, M. C. (1994). Classroom behavior of very low birth weight elementary school children. *Pediatrics, 94*, 700–708.

Kogan, N. (1976). *Cognitive styles in infancy and early childhood*. Hillsdale: Erlbaum.
Laosa, L. M. (1977). Nonbiased assessment of children's abilities: Historical antecedents and current issues. In T. Oakland (Ed.), *Psychological and educational assessment of minority children* (pp. 1–20). New York: Bruner/Mazel.
Laosa, L. M. (1978). Maternal teaching strategies in Chicano families of varied educational and socioeconomic levels. *Child Development, 49*, 1129–1135. https://doi.org/10.2307/1128752
Laosa, L. M. (1980a). Maternal teaching strategies and cognitive styles in Chicano families. *Journal of Educational Psychology, 72*, 45–54. https://doi.org/10.1037/0022-0663.72.1.45
Laosa, L. M. (1980b). Maternal teaching strategies in Chicano and Anglo-American families: The influence of culture and education on maternal behavior. *Child Development, 51*, 759–765. https://doi.org/10.2307/1129462
Laosa, L. M. (1982a). Families as facilitators of children's intellectual development at 3 years of age—A causal analysis. In L. M. Laosa & I. E. Sigel (Eds.), *Families as learning environments for children* (pp. 1–45). New York: Plenum Press. https://doi.org/10.1007/978-1-4684-4172-7_1
Laosa, L. M. (1982b). School, occupation, culture, and family: The impact of parental schooling on the parent–child relationship. *Journal of Educational Psychology, 74*, 791–827. https://doi.org/10.1037/0022-0663.74.6.791
Laosa, L. M. (1982c). The sociocultural context of evaluation. In B. Spodek (Ed.), *Handbook of research in early childhood education* (pp. 501–520). New York: Free Press.
Laosa, L. M. (1984). Ethnic, socioeconomic, and home language influences upon early performance on measures of abilities. *Journal of Educational Psychology, 76*, 1178–1198. https://doi.org/10.1037/0022-0663.76.6.1178
Laosa, L. M. (1990). Psychosocial stress, coping, and development of Hispanic immigrant children. In F. C. Serafica, A. I. Schwebel, R. K. Russell, P. D. Isaac, & L. B. Myers (Eds.), *Mental health of ethnic minorities* (pp. 38–65). New York: Praeger.
Laosa, L. M. (1991). The cultural context of construct validity and the ethics of generalizability. *Early Childhood Research Quarterly, 6*, 313–321. https://doi.org/10.1016/S0885-2006(05)80058-4
Laosa, L. M. (1996). Intelligence testing and social policy. *Journal of Applied Developmental Psychology, 17*, 155–173. https://doi.org/10.1016/S0193-3973(96)90023-4
Laosa, L. M. (1997). Research perspectives on constructs of change: Intercultural migration and developmental transitions. In A. Booth, A. C. Crouter, & N. Landale (Eds.), *Immigration and the family—Research and policy on U.S. Immigrants* (pp. 133–148). Mahwah: Erlbaum.
Laosa, L. M. (1999). Intercultural transitions in human development and education. *Journal of Applied Developmental Psychology, 20*, 355–406. https://doi.org/10.1016/S0193-3973(99)00023-4
Laosa, L. M. (2001). School segregation of children who migrate to the United States from Puerto Rico. *Education Policy Analysis Archives, 9*(1), 1–49. https://doi.org/10.14507/epaa.v9n1.2001
Lee, V. E., Brooks-Gunn, J., & Schnur, E. (1988). Does Head Start work? A 1-year follow-up comparison of disadvantaged children attending Head Start, no preschool, and other preschool programs. *Developmental Psychology, 24*, 210–222. https://doi.org/10.1037/0012-1649.24.2.210
Lee, V. E., Brooks-Gunn, J., Schnur, E., & Liaw, F.-R. (1990). Are Head Start effects sustained? A longitudinal follow-up comparison of disadvantaged children attending Head Start, no preschool, and other preschool programs. *Child Development, 61*, 495–507. https://doi.org/10.2307/1131110
Lewis, M. (1969). Infants' responses to facial stimuli during the first year of life. *Developmental Psychology, 1*, 75–86. https://doi.org/10.1037/h0026995
Lewis, M. (1971a). A developmental study of the cardiac response to stimulus onset and offset during the first year of life. *Psychophysiology, 8*, 689–698. https://doi.org/10.1111/j.1469-8986.1971.tb00506.x
Lewis, M. (1971b). Individual differences in the measurement of early cognitive growth. In J. Hellmuth (Ed.), *Studies in abnormalities: Vol. 2. Exceptional infant* (pp. 172–210). New York: Brunner/Mazel.

Lewis, M. (1971c). Infancy and early childhood in the urban environment: Problems for the 21st century. In European Cultural Foundation (Ed.), *Citizen and city in the year 2000* (pp. 167–172). Deventer: Kluwer.
Lewis, M. (1973). Infant intelligence tests: Their use and misuse. *Human Development, 16*, 108–118. https://doi.org/10.1159/000271270
Lewis, M. (1974). The cardiac response during infancy. In R. F. Thompson (Series Ed.), R. F. Thompson & M. M. Patterson (Vol. Eds.), *Methods in physiological psychology: Vol. 1, Part C. Bioelectric recording techniques: Receptor and effector processes* (pp. 201–229). New York: Academic Press.
Lewis, M. (1975a). Early sex differences in the human: Studies of socioemotional development. *Archives of Sexual Behavior, 4*, 329–335. https://doi.org/10.1007/BF01541719
Lewis, M. (1975b). The development of attention and perception in the infant and young child. In W. M. Cruickshank & D. P. Hallahan (Eds.), *Perceptual and learning disabilities in children* (Vol. 2, pp. 137–162). Syracuse: Syracuse University Press.
Lewis, M. (1976a). A theory of conversation. *Quarterly Newsletter of the Institute for Comparative Human Development, 1*(1), 5–7.
Lewis, M. (1976b). What do we mean when we say "infant intelligence scores"? A sociopolitical question. In M. Lewis (Ed.), *Origins of intelligence: Infancy and early childhood* (pp. 1–18). New York: Plenum Press. https://doi.org/10.1007/978-1-4684-6961-5_1
Lewis, M. (1977a). A new response to stimuli. *The Sciences, 17*(3), 18–19. https://doi.org/10.1002/j.2326-1951.1977.tb01524.x
Lewis, M. (1977b). Early socioemotional development and its relevance to curriculum. *Merrill-Palmer Quarterly, 23*, 279–286.
Lewis, M. (1978a). Attention and verbal labeling behavior in preschool children: A study in the measurement of internal representations. *Journal of Genetic Psychology, 133*, 191–202. https://doi.org/10.1080/00221325.1978.10533377
Lewis, M. (1978b). Situational analysis and the study of behavioral development. In L. A. Pervin & M. Lewis (Eds.), *Perspectives in interactional psychology* (pp. 49–66). New York: Plenum Press. https://doi.org/10.1007/978-1-4613-3997-7_3
Lewis, M. (1980). Developmental theories. In I. L. Kutash & L. B. Schlesinger (Eds.), *Handbook on stress and anxiety—Contemporary knowledge, theory, and treatment* (pp. 48–62). San Francisco: Jossey-Bass.
Lewis, M. (1981a). Attention as a measure of cognitive integrity. In M. Lewis & L. T. Taft (Eds.), *Developmental disabilities—Theory, assessment, and intervention* (pp. 185–212). Jamaica: Spectrum. https://doi.org/10.1007/978-94-011-6314-9_13
Lewis, M. (1981b). Self-knowledge: A social-cognitive perspective on gender identity and sex role development. In M. E. Lamb & L. R. Sherrod (Eds.), *Infant social cognition: Empirical and theoretical considerations* (pp. 395–414). Hillsdale: Erlbaum.
Lewis, M. (1982a). Play as whimsy. *Behavioral and Brain Sciences, 5*, 166. https://doi.org/10.1017/S0140525X00011067
Lewis, M. (1982b). The social network systems model—Toward a theory of social development. In T. M. Field, A. Huston, H. C. Quary, L. Troll, & G. E. Finley (Eds.), *Review of human development* (pp. 180–214). New York: Wiley.
Lewis, M. (n.d.). *My life at ETS.* Unpublished manuscript.
Lewis, M., & Baldini, N. (1979). Attentional processes and individual differences. In G. A. Hale & M. Lewis (Eds.), *Attention and cognitive development* (pp. 137–172). New York: Plenum Press. https://doi.org/10.1007/978-1-4613-2985-5_7
Lewis, M., & Ban, P. (1977). Variance and invariance in the mother–infant interaction: A cross-cultural study. In P. H. Leiderman, S. R. Tulkin, & A. Rosenfeld (Eds.), *Culture and infancy—Variations in the human experience* (pp. 329–355). New York: Academic Press.
Lewis, M., & Baumel, M. H. (1970). A study in the ordering of attention. *Perceptual and Motor Skills, 31*, 979–990. https://doi.org/10.2466/pms.1970.31.3.979

Lewis, M., & Brooks, J. (1974). Self, others and fear: Infants' reactions to people. In M. Lewis & L. A. Rosenblum (Eds.), *The origins of behavior: Vol. 2. The origins of fear* (pp. 195–227). New York: Wiley.

Lewis, M., & Brooks, J. (1975). Infants' social perception: A constructivist view. In L. B. Cohen & P. Salapatek (Eds.), *Perception of space, speech, and sound* (Infant perception: From sensation to cognition, Vol. 2, pp. 101–148). New York: Academic Press.

Lewis, M., & Brooks-Gunn, J. (1979a). *Social cognition and the acquisition of self*. New York: Plenum Press. https://doi.org/10.1007/978-1-4684-3566-5

Lewis, M., & Brooks-Gunn, J. (1979b). The search for the origins of self: Implications for social behavior and intervention. In L. Montada (Ed.), *Brennpunkte der entwicklungspsychologie [Foci of developmental psychology]* (pp. 157–172). Stuttgart: Kohlhammer.

Lewis, M., & Brooks-Gunn, J. (1979c). Toward a theory of social cognition: The development of self. *New Directions for Child Development, 4*, 1–20. https://doi.org/10.1002/cd.23219790403

Lewis, M., & Brooks-Gunn, J. (1981a). Attention and intelligence. *Intelligence, 5*, 231–238. https://doi.org/10.1016/S0160-2896(81)80010-4

Lewis, M., & Brooks-Gunn, J. (1981b). The self as social knowledge. In M. D. Lynch, A. A. Norem-Hebeisen, & K. J. Gergen (Eds.), *Self-concept: Advances in theory and research* (pp. 101–118). Cambridge, MA: Ballinger.

Lewis, M., & Brooks-Gunn, J. (1981c). Visual attention at three months as a predictor of cognitive functioning at two years of age. *Intelligence, 5*, 131–140. https://doi.org/10.1016/0160-2896(81)90003-9

Lewis, M., & Cherry, L. (1977). Social behavior and language acquisition. In M. Lewis & L. A. Rosenblum (Eds.), *The origins of behavior: Vol. 5. Interaction, conversation, and the development of language* (pp. 227–245). New York: Wiley.

Lewis, M., & Feiring, C. (1979). The child's social network: Social object, social functions, and their relationship. In M. Lewis & L. A. Rosenblum (Eds.), *Genesis of behavior: Vol. 2. The child and its family* (pp. 9–27). New York: Plenum Press.

Lewis, M., & Feiring, C. (1981). Direct and indirect interactions in social relationships. *Advances in Infancy Research, 1*, 129–161.

Lewis, M., & Feiring, C. (1982). Some American families at dinner. In L. Laosa & I. Sigel (Eds.), *Families as learning environments for children* (pp. 115–145). New York: Plenum Press. https://doi.org/10.1007/978-1-4684-4172-7_4

Lewis, M., & Fox, N. (1980). Predicting cognitive development from assessments in infancy. *Advances in Behavioral Pediatrics, 1*, 53–67.

Lewis, M., & Freedle, R. (1977). The mother and infant communication system: The effects of poverty. In H. McGurk (Ed.), *Ecological factors in human development* (pp. 205–215). Amsterdam: North-Holland.

Lewis, M., Goldberg, S., & Campbell, H. (1969a). A developmental study of information processing within the first three years of life: Response decrement to a redundant signal. *Monographs of the Society for Research in Child Development, 34*(9, Serial No. 133). https://doi.org/10.2307/1165696

Lewis, M., & McGurk, H. (1973). Evaluation of infant intelligence: Infant intelligence scores—True or false? *Science, 178*(4066), 1174–1177. https://doi.org/10.1126/science.178.4066.1174

Lewis, M., & Michalson, L. (1982a). The measurement of emotional state. In C. E. Izard & P. B. Read (Eds.), *Measuring emotions in infants and children* (Vol. 1, pp. 178–207). New York: Cambridge University Press.

Lewis, M., & Michalson, L. (1982b). The socialization of emotions. In T. Field & A. Fogel (Eds.), *Emotion and early interaction* (pp. 189–212). Hillsdale: Erlbaum.

Lewis, M., & Michalson, L. (1983). *Children's emotions and moods: Developmental theory and measurement*. New York: Plenum Press. https://doi.org/10.1007/978-1-4613-3620-4

Lewis, M., & Rosenblum, L. A. (Eds.). (1974a). *The origins of behavior: Vol. 1. The effect of the infant on its caregiver*. New York: Wiley.

Lewis, M., & Rosenblum, L. A. (Eds.). (1974b). *The origins of behavior: Vol. 2. The origins of fear*. New York: Wiley.
Lewis, M., & Rosenblum, L. A. (Eds.). (1975). *The origins of behavior: Vol. 4. Friendship and peer relations*. New York: Wiley.
Lewis, M., & Rosenblum, L. A. (Eds.). (1977). *The origins of behavior: Vol. 5. Interaction, conversation, and the development of language*. New York: Wiley.
Lewis, M., & Rosenblum, L. A. (Eds.). (1978a). *Genesis of behavior: Vol. 1. The development of affect*. New York: Plenum Press.
Lewis, M., & Rosenblum, L. A. (1978b). Introduction: Issues in affect development. In M. Lewis & L. A. Rosenblum (Eds.), *Genesis of behavior: Vol. 1. The development of affect* (pp. 1–10). New York: Plenum Press.
Lewis, M., & Rosenblum, L. A. (Eds.). (1979). *Genesis of behavior: Vol. 2. The child and its family*. New York: Plenum Press.
Lewis, M., & Rosenblum, L. A. (Eds.). (1981). *Genesis of behavior: Vol. 3. The uncommon child.* New York: Plenum Press.
Lewis, M., & Schaeffer, S. (1981). Peer behavior and mother–infant interaction in maltreated children. In M. Lewis & L. A. Rosenblum (Eds.), *Genesis of behavior: Vol. 3. The uncommon child* (pp. 193–223). New York: Plenum Press. https://doi.org/10.1007/978-1-4684-3773-7_9
Lewis, M., & Spaulding, S. J. (1967). Differential cardiac response to visual and auditory stimulation in the young child. *Psychophysiology, 3*, 229–237. https://doi.org/10.1111/j.1469-8986.1967.tb02700.x
Lewis, M., & Starr, M. D. (1979). Developmental continuity. In J. D. Osofsky (Ed.), *Handbook of infant development* (pp. 653–670). New York: Wiley.
Lewis, M., & Taft, L. T. (Eds.). (1982). *Developmental disabilities: Theory, assessment, and intervention*. Jamaica: Spectrum. https://doi.org/10.1007/978-94-011-6314-9
Lewis, M., & Wehren, A. (1982). The central tendency in study of the handicapped child. In D. Bricker (Ed.), *Intervention with at-risk and handicapped infants – From research to application* (pp. 77–89). Baltimore: University Park Press.
Lewis, M., & Weinraub, M. (1974). Sex of parent × sex of child: Socioemotional development. In R. C. Friedman, R. M. Richart, & R. L. Vande Wiele (Eds.), *Sex differences in behavior* (pp. 165–189). New York: Wiley.
Lewis, M., & Wilson, C. D. (1970). The cardiac response to a perceptual cognitive task in the young child. *Psychophysiology, 6*, 411–420. https://doi.org/10.1111/j.1469-8986.1970.tb01751.x
Lewis, M., & Zarin-Ackerman, J. (1977). Early infant development. In R. E. Behrman, J. M. Driscoll Jr., & A. E. Seeds (Eds.), *Neonatal-perinatal medicine—Diseases of the fetus and infant* (2nd ed., pp. 195–208). St. Louis: Mosby.
Lewis, M., Kagan, J., Campbell, H., & Kalafat, J. (1966a). The cardiac response as a correlate of attention in infants. *Child Development, 37*, 63–71. https://doi.org/10.2307/1126429
Lewis, M., Kagan, J., & Kalafat, J. (1966b). Patterns of fixation in the young infant. *Child Development, 37*, 331–341. https://doi.org/10.2307/1126807
Lewis, M., Bartels, B., Campbell, H., & Goldberg, S. (1967a). Individual differences in attention – The relation between infants' condition at birth and attention distribution within the first year. *American Journal of Diseases of Children, 113*, 461–465. https://doi.org/10.1001/archpedi.1967.02090190107010
Lewis, M., Goldberg, S., & Rausch, M. (1967b). Attention distribution as a function of novelty and familiarity. *Psychonomic Science, 7*, 227–228. https://doi.org/10.3758/BF03328553
Lewis, M., Dodd, C., & Harwitz, M. (1969b). Cardiac responsivity to tactile stimulation in waking and sleeping infants. *Perceptual and Motor Skills, 29*, 259–269. https://doi.org/10.2466/pms.1969.29.1.259
Lewis, M., Wilson, C. D., Ban, P., & Baumel, M. H. (1970). An exploratory study of resting cardiac rate and variability from the last trimester of prenatal life through the first year of postnatal life. *Child Development, 41*, 799–811. https://doi.org/10.2307/1127225
Lewis, M., Wilson, C. D., & Baumel, M. (1971). Attention distribution in the 24-month-old child: Variations in complexity and incongruity of the human form. *Child Development, 42*, 429–438.

Lewis, M., Young, G., Brooks, J., & Michalson, L. (1975). The beginning of friendship. In M. Lewis & L. Rosenblum (Eds.), *The origins of behavior: Vol. 4. Friendship and peer relations* (pp. 27–66). New York: Wiley.

Lewis, M., Brooks, J., & Haviland, J. (1978). Hearts and faces: A study in the measurement of emotion. In M. Lewis & L. A. Rosenblum (Eds.), *Genesis of behavior: Vol. 1. The development of affect* (pp. 77–123). New York: Plenum Press. https://doi.org/10.1007/978-1-4684-2616-8_4

Lewis, M., Brooks-Gunn, J., & Jaskir, J. (1985). Individual differences in visual self-recognition as a function of mother–infant attachment relationship. *Developmental Psychology, 21*, 1181–1187. https://doi.org/10.1037/0012-1649.21.6.1181

Linn, R. L. (1989). Current perspectives and future directions. In R. L. Linn (Ed.), *Educational measurement* (3rd ed., pp. 1–10). New York: Macmillan.

McCarthy, D. (1972). *Manual for the McCarthy Scales of Children's Abilities*. New York: Psychological Corporation.

McCormick, M. C., Brooks-Gunn, J., Shorter, T., Wallace, C. Y., Holmes, J. H., & Heagarty, M. C. (1987). The planning of pregnancy among low-income women in central Harlem. *American Journal of Obstetrics and Gynecology, 156*, 145–149. https://doi.org/10.1016/0002-9378(87)90226-2

McCormick, M. C., Brooks-Gunn, J., Shorter, T., Holmes, J. H., & Heagarty, M. C. (1989a). Factors associated with maternal rating of infant health in central Harlem. *Journal of Developmental & Behavioral Pediatrics, 10*, 139–144. https://doi.org/10.1097/00004703-198906000-00004

McCormick, M. C., Brooks-Gunn, J., Shorter, T., Holmes, J. H., Wallace, C. Y., & Heagarty, M. C. (1989b). Outreach as case finding: Its effect on enrollment in prenatal care. *Medical Care, 27*, 103–111. https://doi.org/10.1097/00005650-198902000-00002

McCormick, M. C., Brooks-Gunn, J., Workman-Daniels, K., Turner, J., & Peckham, G. J. (1992). The health and developmental status of very low-birth-weight children at school age. *JAMA, 267*, 2204–2208. https://doi.org/10.1001/jama.1992.03480160062035

McGillicuddy-DeLisi, A. V. (1982). The relationship between parents' beliefs about development and family constellation, socioeconomic status, and parents' teaching strategies. In L. M. Laosa & I. E. Sigel (Eds.), *Families as learning environments for children* (pp. 261–299). New York: Plenum Press. https://doi.org/10.1007/978-1-4684-4172-7_9

McGillicuddy-DeLisi, A. V., Sigel, I. E., & Johnson, J. E. (1979). The family as a system of mutual influences: Parental beliefs, distancing behaviors, and children's representational thinking. In M. Lewis & L. A. Rosenblum (Eds.), *Genesis of behavior: Vol. 2. The child and its family* (pp. 91–106). New York: Plenum Press.

McGurk, H., & Lewis, M. (1974). Space perception in early infancy: Perception within a common auditory-visual space? *Science, 186*(4164), 649–650. https://doi.org/10.1126/science.186.4164.649

Messick, S. (1989). Validity. In R. L. Linn (Ed.), *Educational measurement* (3rd ed., pp. 13–103). New York: Macmillan.

Oakland, T., & Laosa, L. M. (1977). Professional, legislative, and judicial influences on psychoeducational assessment practices in schools. In T. Oakland (Ed.), *Psychological and educational assessment of minority children* (pp. 21–51). New York: Bruner/Mazel.

Pellegrini, A. D., Brody, G. H., & Sigel, I. E. (1985). Parents' book-reading habits with their children. *Journal of Educational Psychology, 77*, 332–340. https://doi.org/10.1037/0022-0663.77.3.332

Pellegrini, A. D., Brody, G. H., McGillicuddy-DeLisi, A. V., & Sigel, I. E. (1986). The effects of children's communicative status and task on parents' teaching strategies. *Contemporary Educational Psychology, 11*, 240–252. https://doi.org/10.1016/0361-476X(86)90020-2

Renninger, K. A. (2007). Irving E. (Irv) Sigel (1921–2006). *American Psychologist, 62*, 321. https://doi.org/10.1037/0003-066X.62.4.321

Ruble, D. N., & Brooks-Gunn, J. (1979a). Menstrual myths. *Medical Aspects of Human Sexuality, 13*, 110–121.

Ruble, D. N., & Brooks-Gunn, J. (1979b). Menstrual symptoms: A social cognitive analysis. *Journal of Behavioral Medicine, 2*, 171–194. https://doi.org/10.1007/BF00846665

Ruble, D. N., & Brooks-Gunn, J. (1982). The experience of menarche. *Child Development, 53*, 1557–1566. https://doi.org/10.2307/1130084

Ruble, D. N., Boggiano, A. K., & Brooks-Gunn, J. (1982). Men's and women's evaluations of menstrual-related excuses. *Sex Roles, 8*, 625–638. https://doi.org/10.1007/BF00289896

Schnur, E., Brooks-Gunn, J., & Shipman, V. C. (1992). Who attends programs serving poor children? The case of Head Start attendees and non-attendees. *Journal of Applied Developmental Psychology, 13*, 405–421. https://doi.org/10.1016/0193-3973(92)90038-J

Shipman, V. C. (1972). Introduction. In V. C. Shipman (Ed.), *Disadvantaged children and their first school experiences: ETS–Head Start longitudinal study —Technical reports series* (Program Report No. 72-27, pp. 1–23). Princeton: Educational Testing Service.

Shipman, V. C. (1973). *Disadvantaged children and their first school experiences: ETS–Head Start longitudinal study—Interim report* (Program Report No. 73–35). Princeton: Educational Testing Service.

Sigel, I. E. (1982). The relationship between parental distancing strategies and the child's cognitive behavior. In L. M. Laosa & I. E. Sigel (Eds.), *Families as learning environments for children* (pp. 47–86). New York: Plenum Press. https://doi.org/10.1007/978-1-4684-4172-7_2

Sigel, I. E. (1986). Early social experience and the development of representational competence. *New Directions for Child Development, 1986*(32), 49–65. https://doi.org/10.1002/cd.23219863205

Sigel, I. E. (1992). The belief–behavior connection: A resolvable dilemma? In I. E. Sigel, A. V. McGillicuddy-DeLisi, & J. J. Goodnow (Eds.), *Parental belief systems: The psychological consequences for children* (2nd ed., pp. 433–456). Hillsdale: Erlbaum.

Sigel, I. E. (1993). The centrality of a distancing model for the development of representational competence. In R. R. Cocking & K. A. Renninger (Eds.), *The development and meaning of psychological distance* (pp. 141–158). Hillsdale: Erlbaum.

Sigel, I. E. (2006). Research to practice redefined. In W. Damon & R. M. Lerner (Series Eds.), K. A. Renninger & I. E. Sigel (Vol. Eds.), *Handbook of child psychology: Vol. 4. Child psychology in practice* (6th ed., pp. 1017–1023). New York: Wiley.

Sigel, I. E., & McGillicuddy-DeLisi, A. V. (1984). Parents as teachers of their children: A distancing behavior model. In A. D. Pellegrini & T. D. Yawkey (Eds.), *The development of oral and written language in social contexts* (pp. 71–92). Norwood: Ablex.

Sigel, I. E., & Olmsted, P. (1970). The modification of cognitive skills among lower-class Black children. In J. Hellsmuth (Ed.), *The disadvantaged child: Vol. 3. Compensatory education—A national debate* (pp. 330–338). New York: Bruner/Mazel.

Sigel, I. E., & Saunders, R. A. (1983). On becoming a thinker: An educational preschool program. *Early Child Development and Care, 12*, 39–65. https://doi.org/10.1080/0300443830120105

Sigel, I. E., Flaugher, J., Johnson, J. E., & Schauer, A. (1977). *Manual, Parent–Child Interaction Observation Schedule*. Princeton: Educational Testing Service.

Sontag, L. W., Steele, W. G., & Lewis, M. (1969). The fetal and maternal cardiac response to environmental stress. *Human Development, 12*, 1–9. https://doi.org/10.1159/000270674

Steele, W. G., & Lewis, M. (1968). A longitudinal study of the cardiac response during a problem-solving task and its relationship to general cognitive function. *Psychonomic Science, 11*, 275–276. https://doi.org/10.3758/BF03328189

Suárez-Orozco, C., & Carhill, A. (2008). Afterword: New directions in research with immigrant families and their children. *New Directions for Child and Adolescent Development, 121*, 87–104. https://doi.org/10.1002/cd.224

Suárez-Orozco, C., Suárez-Orozco, M. M., & Todorova, I. (2008). *Learning a new land: Immigrant students in American society*. Cambridge, MA: Harvard University Press.

Swenson, C. H., & Fiore, A. (1975). Scale of Marriage Problems. *1975 Annual Handbook for Group Facilitators, 7,* 72–89.

Terman, L. M., & Merrill, M. A. (1960). *Manual, Stanford Binet Intelligence Scale* (3rd ed.). Boston: Houghton Mifflin.

Thurman, S. K., & Lewis, M. (1979). Children's response to differences: Some possible implications for mainstreaming. *Exceptional Children, 45*, 468–470.
Vygotsky, L. (1978). *Mind in society – The development of higher psychological processes*. Cambridge, MA: Harvard University Press.
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.
Ward, W. C. (1968). Creativity in young children. *Child Development, 39*, 737–754. https://doi.org/10.2307/1126980
Warren, M. P., Brooks-Gunn, J., Hamilton, L. H., Warren, L. F., & Hamilton, W. G. (1986). Scoliosis and fractures in young ballet dancers – Relation to delayed menarche and secondary amenorrhea. *New England Journal of Medicine, 314*, 1348–1353. https://doi.org/10.1056/NEJM198605223142104
Warren, M. P., Brooks-Gunn, J., Fox, R. P., Lancelot, C., Newman, D., & Hamilton, W. G. (1991). Lack of bone accretion and amenorrhea: Evidence for a relative osteopenia in weight-bearing bones. *Journal of Clinical Endocrinology & Metabolism, 72*, 847–853. https://doi.org/10.1210/jcem-72-4-847
Wechsler, D. B. (1949a). *Wechsler Intelligence Scale for Children manual*. New York: Psychological Corporation.
Wechsler, D. (1949b). *Wechsler Preschool and Primary School Scale of Intelligence*. New York: Psychological Corporation.
Wechsler, D. B. (1955). *Manual for the Wechsler Adult Intelligence Scale*. New York: Psychological Corporation.
Weinraub, M., & Lewis, M. (1977). The determinants of children's responses to separation. *Monographs of the Society for Research in Child Development, 42*(4, Serial No. 172). https://doi.org/10.2307/1165949
Weinraub, M., Brooks, J., & Lewis, M. (1977). The social network: A reconsideration of the concept of attachment. *Human Development, 20*, 31–47. https://doi.org/10.1159/000271546
Wertsch, J. V., McNamee, G. D., McLane, J. B., & Budwig, N. A. (1980). The adult–child dyad as a problem solving system. *Child Development, 51*, 1215–1221. https://doi.org/10.2307/1129563
Williams, R. L. (1972). *Manual of directions, Black Intelligence Test of Cultural Homogeneity*. St. Louis: Author.
Witkin, H. A., Oltman, P. K., Raskin, E., & Karp, S. A. (1971). *A manual for the Embedded Figures Tests*. Palo Alto: Consulting Psychologists Press.
Zarin-Ackerman, J., Lewis, M., & Driscoll Jr., J. M. (1977). Language development in 2-year-old normal and risk infants. *Pediatrics, 59*, 982–986.