

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

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-shalowness 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 \times viewpoint \times high vs. low quality) contingency table was constructed, and its chi-square was partitioned to yield main and interaction effects. A highly significant viewpoint \times 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 (reabeled 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; Zocolotti 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 intra-group 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 augments-reduces 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-aftereffect 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 augments-reducers 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 postinduction 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 augments-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 augments-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 augments-reduces 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 augments-reduces 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 augementer-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 augementer-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](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](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](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 intellectual 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 kinesthesia. 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](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](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. Cognitive 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/s15326985Sep2903_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](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](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](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](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](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 configural processing. *Cortex*, 14, 155–163. [https://doi.org/10.1016/S0010-9452\(78\)80040-9](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.

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

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

