Social Psychology of Creativity: A Consensual Assessment Technique

Teresa M. Amabile
Brandeis University

Although personality research has made much progress in developing an individual-difference psychology of creativity, the nature of this phenomenon can only be fully illuminated if a social psychology of creativity is developed as well. Yet experimental studies of social and environmental influences on creativity are extremely rare. A major obstacle to such research is the criterion problem: the lack of a clear operational definition and an appropriate assessment methodology. For a variety of reasons, both the popular creativity tests and the subjective assessment techniques used in some previous creativity studies are ill-suited to social psychological studies of creativity. A consensual definition of creativity is presented and, as a refinement of previous subjective methods, a reliable subjective assessment technique based on that definition is described. Several studies testing the methodology in both artistic and verbal domains are presented, and the advantages and limitations of this technique are discussed. The present methodology can be particularly useful for the development of a social psychology of creativity because of the nature of the tasks employed and the nature of the creativity assessments obtained. In the discussion, creativity assessment is considered as one instance illustrating the more general issue of the divergent aims and methods of personality psychology and social psychology.

The past 30 years have seen an enormous increase in creativity research. Creativity citations, which accounted for less than .2% of the listings in *Psychological Abstracts* in 1950, made up more than 1% of the citations in 1980. Despite this proliferation of research publications on creativity, however, there has been virtually no work on the social psychology of creativity. Between 1975 and 1980, there were barely half a dozen articles in the *Journal of Personality and Social Psychology* and the *Journal of Experimental Social Psychology* that dealt in some way with the effects of social and environmental factors on creativity. Instead, creativity research has been heavily personality oriented; studies on the description and identification of creative personality and cognitive style have predominated (e.g., Gough, 1979; Lott, 1978; Sobel & Rothenberg, 1980).

Development of a social psychology of creativity could be important for both personality and social psychology, as well as for the study of creativity. Although a great deal of research has already focused on the creative personality, an illumination of the types of social variables that can positively or negatively influence creative performance would aid in the comprehensive description of cre-

This work was supported in part by Predoctoral Fellowship 5F31 MH05232-02 from the National Institute of Mental Health and by a BRSG SO7044 awarded by the Biomedical Research Support Grant Program, Division of Research Resources, the National Institutes of Health.

The author gratefully acknowledges the help of several colleagues, students, and friends who worked on the studies reported here: Carolyn Amabile, Phyllis Amabile, and Julia Steinmetz, who provided invaluable assistance in the design and execution of Study 1; Elliot Eisner, who offered several suggestions for Study 1; Steven Berglas and Marie Handel, who collaborated on Study 2; Margaret Stubbs, who collaborated on Study 3; Shereen Brackfield, who collaborated on Study 6; Phyllis Goldfarb, who collaborated on Study 7; Lisa Berman and Ronit Goldlust who helped conduct Study 8; and the principals, teachers, and students of St. Jude's School in Waltham, Massachusetts, and Charles E. Cashman School in Amesbury, Massachusetts, for their invaluable cooperation in Studies 2 and 3.

In addition, several colleagues offered very helpful comments on an earlier draft of this manuscript: William DeJong, Maurice Hershenson, Robert Kidd, Ellen Langer, Mark Lepper, Leslie McArthur, Ricardo Morant, Harvey Pines, Lee Ross, Mark Snyder, and Margaret Stubbs.

Requests for reprints should be sent to Teresa M. Amabile, Department of Psychology, Brandeis University, Waltham, Massachusetts, 02254.

Journal of Personality and Social Psychology, 1982, Vol. 43, No. 5, 997-1013
 0022-3514/82/4305-0997$00.75

ative personality, its development, and its manifestation. Within social psychology, researchers have in recent years turned increasing attention to an exploration of the effects of social factors on various aspects of cognitive and motor performance (e.g., Condry, 1977; Deci, 1971; Diener & Dweck, 1978; Kruglanski, Friedman, & Zeevi, 1971; Lepper, Greene, & Nisbett, 1973; McGraw & McCullers, 1979; Seta & Hassan, 1980; Wimperis & Farr, 1979). As one important aspect of performance that has been virtually ignored in these investigations, creativity should be integrated into social psychological theories of motivation and cognition. Finally, although cognitive and characterological determinants of creativity unquestionably deserve research attention, theories of the creative process will be incomplete without an accounting of social-situational determinants as well.

Methodologically, studies of the social psychology of creativity can be grouped into two broadly defined categories: archival and experimental. An excellent example of the former type is a program of research carried out over the past few years by Simonton (1975, 1977a, 1977b, 1979). In these studies, data collected from archival sources such as histories, anthologies, and biographical dictionaries are subjected to transhistorical time-series analyses to test hypotheses about the effect of social variables such as role model availability and political instability on creative production. Despite the utility of this approach, however, it must be coupled with experimental techniques to more clearly identify the causal influence of social factors on creativity and the mechanism of their operation (Simonton, 1975). It is such experimental investigations that are so rare in the literature of personality and social psychology.

Most of the experimental work on creativity has used one of two assessment techniques. Some researchers have had judges rate products (or the persons who produced them) on scales of creativity. The vast majority, however, have relied upon creativity tests—instruments that are usually similar in form and administration to conventional intelligence tests. Despite these two rather well-established approaches to the assessment of creativity, it can be argued that the criterion problem still presents a major obstacle to successful and rigorous empirical work in all areas of creativity research and especially in the social psychology of creativity. The definition and assessment of creativity have long been the subject of disagreement and dissatisfaction among psychologists. This problem has been dealt with in a variety of ways by the few researchers who have experimentally studied social influences on creativity, but none of them has explicitly addressed the adequacy of conventional assessment methods in any detail.

Creativity Tests

Although several different batteries of creativity tests have been developed, the most widely used (and the prototype for many other tests) are the Torrance Tests of Creative Thinking (TTCT; Torrance, 1962). The TTCT consist of a number of subtests, most of which are verbal, although some involve drawing. The child (or adult, depending on the form used) is typically presented with one or several subtests, such as the Unusual Uses test, in which the child is asked to think of as many different uses as he can for a common object; the Consequences Test, in which the child is required to suggest consequences of several "impossible" events (e.g., people becoming invisible at will); and Imaginative Stories, where the child is asked to write an interesting and exciting story on a given topic (e.g., the dog that doesn't bark). Responses to these test items are scored according to the number and variety of ideas expressed, as well as the elaboration of ideas and their statistical infrequency. Most other creativity tests are similar to the TTCT in form, content, administration, and scoring (e.g., Guilford, 1967; Wallach & Kogan, 1965).

One conceptual problem with the use of creativity tests in empirical studies of creative performance is that their use is not clearly tied to an operational definition of creativity. There are, certainly, many definitions of creativity. Although some theorists have expressed the bold hope that an ultimate definition will one day be articulated (Ghiselin, 1963), it is generally recognized that a search for specific, objectively identifiable features

common to all creative products will not be fruitful. Instead, most definitions, while using the creative product as the distinguishing sign of creativity, propose that the general qualities of novelty and appropriateness differentiate creative from uncreative products (e.g., Barron, 1955; Bruner, 1962; Newell, Shaw, & Simon, 1962; Stein, 1974). In other words, the product or response must be unusual, statistically infrequent, or completely unique, and it must also be correct in the context of the problem or audience to which it was addressed. In addition, some theorists define creativity in terms of the response that the product produces in an observer. For example, Bruner (1962) saw the creative product as anything that produces "effective surprise" in the observer, along with a "shock of recognition" that the product or response, although novel, is entirely appropriate.

Despite the existence of such intuitively reasonable definitions, however, current assessment techniques depending on the creativity tests are not closely linked to them. Indeed, it can be argued that the criterion problem in creativity research has arisen in large part because most definitions do not include conceptualizations that are readily translated into useful assessment criteria, let alone ultimate criteria. It is not clear, for example, how one might assess novelty and appropriateness in any general way or how one might determine in any satisfactory way whether an observer is experiencing "effective surprise" and a "shock of recognition." As a result, the creativity tests and the studies built around them are operating in a definitional void. There is no clear, explicit statement of the criteria that conceptually underlie the assessment procedures.

A second problem is the basically subjective nature of the purportedly objective scoring procedures in the creativity tests. Performance on these tests is rated according to the test constructor's or scorer's intuitive assessment of what is creative and not according to *objective* criteria of novelty, appropriateness, satisfyingness, and so on. Motivated perhaps by the apparent success of objective tests of intelligence, creativity researchers might have been too quick to attempt to objectify the assessment of creativity. Almost certainly, many creativity tests do measure abilities that are important for creative performance, but it is unclear whether they are useful for directly assessing something called *creativity*. It will be argued here that such judgments can ultimately only be subjective.

For the purposes of developing a social psychology of creativity, however, the most crucial feature of the creativity tests is that they were primarily developed as tools for personality research. They were expressly designed to be sensitive to individual differences in performance in a wide variety of domains—verbal, artistic, scientific, and mathematical (e.g., Guilford, 1956, 1957a, 1957b; Torrance, 1962). To the extent that they are able to detect subtle individual differences, however, these tests are problematic for experimental studies of social and environmental influences on creativity. Although it has been suggested that creativity tests can be influenced by social and environmental factors (e.g., Wallach & Kogan, 1965), there is abundant evidence that many of these tests do assess rather stable individual differences (e.g., Gakhar & Luthra, 1973; Holland, 1968; Torrance, 1972a, 1972b). It is precisely this success in identifying stable traits that renders creativity tests less appropriate for studies of environmental influences (Feldman, 1980). Normally, social psychologists seek to control for and, as much as possible, eliminate individual-difference or within-group variability in the crucial dependent measures so that those measures can more easily detect the "signal" of between-group differences produced by experimental manipulations (Carlsmith, Ellsworth, & Aronson, 1976). Thus, in developing a social psychology of creativity, it would seem unwise to rely upon assessment techniques that were expressly designed to reveal consistent individual differences.

Judgments of Products

The second, and much less common, approach to creativity assessment has been the use of judges to provide subjective ratings of the creativity of products or the persons who produced them (e.g., Domino, 1974; Helson & Crutchfield, 1970; Kruglanski et al., 1971; MacKinnon, 1962; Sobel & Rothenberg, 1980). In one of the best known of the studies

using this approach, for example, MacKinnon and his colleagues (1962) asked five university professors of architecture to nominate the 40 most creative architects in the United States and to rate the creativity of each on a 5-point scale. The nominated architects were then invited to participate in extensive personality assessments. In other studies (e.g., Sobel & Rothenberg, 1980), subjects with special expertise or experience in a particular domain are asked to produce works in that domain during the experimental session; these products are then rated on creativity by expert judges.

Studies that have employed subjective judgment have clearly avoided one of the problems noted earlier with the creativity tests. In their use of judges' ratings, they present an assessment of creativity in which the subjective nature of the measure is direct and unveiled, in contrast to the apparent objectivity of the creativity tests. In other respects, however, even these methods present difficulties. These studies, too, often suffer from a definitional void by failing to explictly articulate the definition of creativity guiding the research or by presenting conceptual definitions (usually including novelty and appropriateness as criteria) that are not directly tied to the assessment procedure. In addition, in their assessment procedures, these studies often fail to differentiate between the creativity of the products and other constructs, such as technical correctness or aesthetic appeal. Moreover, the interjudge reliability might be questioned in studies where the experimenter presents judges with his or her own definition of creativity for them to apply or trains judges beforehand to agree with one another (e.g., Eisner, 1965).

Of primary importance for the present discussion, however, are the ways in which previous subjective methodologies might be inappropriate for use in social psychological research. Although these techniques may not have been specifically designed for personality research, those that require judges to make assessments of a person on the basis of a large body of work over time are likely to detect relatively stable characteristics and would, therefore, certainly be better suited to personality than to social-psychological research.

Even those assessment procedures that have judges rate a single product, however, may be too sensitive to large and stable individual differences in performance. In a recent conceptualization of the creative process (Amabile, in press), it was proposed that creativity depends on three classes of factors: domain-relevant skills, such as knowledge about and talent in the domain, which depend on innate abilities and training; creativity-relevant skills, such as cognitive style, work habits, and knowledge of creativity heuristics, which depend on personality characteristics and training; and task motivation, which depends on individual task preferences and socially imposed constraints. A social-psychological study might, for example, investigate decrements in creativity resulting from a decrease in task motivation by the offer of a tangible reward. The assessment of creativity will be most sensitive to such task motivational effects if the influence of domain-relevant and creativity-relevant skills can be controlled or eliminated. Thus, to the extent that the task presented to subjects draws on special talents or experience-related skills—as the tasks in most previous subjective-assessment studies do—the assessment will be insensitive to social-psychological effects. Moreover, the high levels of skill required by some tasks employed will limit possible subject populations to those who have considerable experience in that domain.

It appears, then, that the criterion problem presents a major obstacle to the investigation of social-environmental effects on creativity. The development of a social psychology of creativity that could be integrated with personality and cognitive research might be facilitated by a straightforward operational definition of creativity and by an assessment technique amenable to social-psychological inquiry. The definition and the methodology described here are presented to that end.

A Consensual Definition of Creativity

The creativity assessment technique described here is grounded in a consensual definition of creativity—an explicitly operational definition that implicitly underlies most subjective creativity assessment methodologies:

A product or response is creative to the extent that appropriate observers independently agree it is creative. Appropriate observers are those familiar with the domain in which the product was created or the response articulated. Thus, creativity can be regarded as the quality of products or responses judged to be creative by appropriate observers, and it can also be regarded as the process by which something so judged is produced.

It should be noted, first, that this definition, like most current definitions of creativity, is based on the creative *product* rather than on the process or the person. Not only is a clear articulation of the creative process not yet possible but, more importantly, any identification of a thought process as creative must finally depend on the fruit of that process—a product or response. Similarly, even a clear specification of the personality traits that mark outstandingly creative individuals would have to be validated against their work. Thus, a product-centered operational definition is clearly most useful for empirical research in creativity.

Perhaps the most important feature of this definition is its reliance on subjective criteria. In this way, it overcomes the difficulty of attempting to specify ultimate objective criteria for identifying products as creative. Indeed, it can be argued that it is impossible to articulate such ultimate criteria—that, just as the judgment of attitude statements as more or less favorable (Thurstone & Chave, 1929) or the identification of individuals as physically attractive (Walster, Aronson, Abrahams, & Rottman, 1966) depends on a social context, so too does the judgment of creativity. Certainly, there must be particular *characteristics* of attitude statements of persons or products that observers systematically look to in rating them on scales of favorability or physical attractiveness or creativity, but, ultimately, the choice of those characteristics is a subjective one. This assertion of the necessarily subjective nature of creativity assessment must be emphatically underscored, given the strong predominance of creativity tests in empirical research and the tendency for many observers to regard these tests as pure, objective measures of creativity.

This consensual definition of creativity also avoids the translation problem noted earlier with traditional definitions. Although the conceptualizations of novelty and appropriateness in traditional definitions are undoubtedly important in formulating theoretical models of creative performance (see Amabile, in press), it is not clear how these concepts are to be translated into assessment operations. By contrast, the consensual definition conceptually identifies creativity with the assessment operations. It may indeed be possible to identify particular objective features of products that correlate with subjective judgments of creativity or to analyze the nature of subjective correlates of those judgments, but this definition makes it unnecessary to attempt to specify those objective features or the characteristics of those subjective reactions beforehand.

The consensual definition rests on a pair of important assumptions. The first assumption is that it is possible to obtain reliable judgments of the creativity of products, given an appropriate group of judges. In other words, although creativity in a product may be difficult to characterize in terms of specific features, it is something that people can recognize when they see it. Furthermore, people familiar with such products can agree with one another on this perception. Second, the assessment methodology that is built upon the consensual definition assumes that there are degrees of creativity—that observers can indeed say, at an acceptable level of agreement, that some products are more creative or less creative than others. This assumption of a continuous underlying dimension of creativity is fairly standard in contemporary theory and research on creativity (e.g., Getzels & Jackson, 1962; Guilford, 1959; Nicholls, 1972).

A Reliable Subjective Assessment Technique

Given the methodological requirements of a social psychology of creativity and given the consensual definition and the assumptions about creativity asserted here, there are several features required of an appropriate subjective assessment methodology. First, there are some requirements of the task itself. Because it is desirable for social-psychological research that there not be large individual differences in baseline performances on the

task, it should be one that does not depend heavily on certain specialized skills—such as drawing ability or verbal facility—that some individuals have undoubtedly developed more fully than others. At the same time, the task should be open-ended enough to permit considerable flexibility in responses. It is desirable that the task allow for novelty in response, since the independent variables will be designed to influence creativity and since, conceptually, creativity has most often been identified with novelty of response. In addition, of course, the task must lead to some product or clearly observable response that can be made available to appropriate judges for assessment.

There are also a number of requirements for the assessment procedure. First, the judges should all have some experience with the domain in question, although the level of experience for all judges need not be identical. Basically, the method requires that all judges be familiar enough with the domain to have developed, over a period of time, some implicit criteria for creativity, technical goodness, and so on. It is important to note that these judges need not themselves have produced work rated as highly creative; it is their familiarity with the domain that is important.

Second, the judges must make their assessments independently. Because the technique relies on subjective judgments, particularly the implicity shared subjective standards of creativity that judges are proposed to have, these judges should not be trained by the experimenter to agree with one another, and they should not be given specific criteria for judging creativity.

Third, in preliminary work on developing the technique for a given task, judges should be asked to make assessments on other dimensions in addition to creativity. Minimally, they should make ratings of the technical aspects of the work and, if appropriate, its aesthetic appeal as well. This would then make it possible to examine the degree of relatedness or independence of those dimensions in subjective assessments of the product in question. Assessments of other aspects of the work would also allow for an analysis of effects of social and environmental factors on those aspects, making a comparison with creativity effects possible. This is important because theoretically there might be reasons to predict a differential effect of a given social factor on creativity and on technical performance.

Fourth, judges should be instructed to rate the products relative to one another on the dimensions in question rather than rating them against some absolute standards they might have for sculpture, poetry, and so on. This is important because for most studies the levels of creativity produced by the "ordinary" subjects who participate will be very low in comparison with the greatest works ever produced in that domain.

Finally, each judge should view the products in a different random order and each judge should consider the various dimensions of judgment in a different random order. If all judgments were made in the same order by all judges, high levels of interjudge reliability might reflect method artifacts.

Once the judgments are obtained, ratings on each dimension should be analyzed for interjudge reliability. In addition, if several subjective dimensions of judgment have been obtained, these should be subjected to a factor analysis to determine the degree of independence between creativity and the other dimensions. Finally, if the products lend themselves to a straightforward identification of specific objective features, these features may be assessed and correlated with creativity judgments.

Given the consensual definition of creativity, the most important criterion for the results of this assessment procedure is that the ratings be reliable. By definition, interjudge reliability in this method is equivalent to construct validity; if appropriate judges independently agree that a given product is highly creative, then it can and must be accepted as such. In addition, it should be possible to separate subjective judgments of creativity from judgments of technical goodness and from judgments of aesthetic appeal. Obviously, for some domains of endeavor, it may be relatively difficult to obtain ratings of aesthetic appeal and technical quality that are not highly correlated with ratings of creativity. However, it is important to demonstrate that it is at least possible to separate these dimensions. Otherwise, the discrimi-

nant validity of the measure would be in doubt; judges might be rating something as creative simply because they like it or because they find it to be technically well done.

Judges' ratings can be used to determine if the original task presented to subjects was appropriate for the purposes of a social-psychological methodology. Certainly, if virtually all of the subjects in a random sample of a population are able to do the task and report no technical difficulty in doing so (i.e., in manipulating the materials, in finishing within a reasonable period of time, and so on), this suggests that the task was well-chosen for these purposes. If later judging of the products reveals a low correlation between judged creativity and experience-related characteristics of the subjects (e.g., age, experience with the particular type of materials), then the task can truly be considered a satisfactory one.[1]

Application to Artistic Creativity

In a program of research carried out over the last 5 years, several studies have demonstrated that it is possible to develop a reliable subjective assessment technique appropriate for social-psychological studies of creativity. Most of these examined artistic creativity, an area that has often been neglected by traditional assessment methods.

Study 1

Method

Subjects and judges. Subjects were 22 girls, aged 7–11 years, who each participated in one of two group sessions held in their apartment complex in Palo Alto, California. Because this was the preliminary investigation of this assessment technique, three different groups of judges were used, varying in expertise with art: (a) psychologist judges—12 members of the Stanford University psychology department (faculty and graduate students), (b) art teachers—21 members of an art education course in the Stanford University School of Education (many of whom were elementary- and secondary-school art teachers), (c) artist judges—7 undergraduate and graduate artists from the art department at Stanford University, each of whom had spent at least 5 years working in studio art.[2]

Materials. All subjects were given identical sets of materials to work with: over 100 pieces of lightweight paper in several different sizes, shapes, and colors (all arranged identically for each subject), a container of glue, and a 15 inch × 20 inch (.38 m × .51 m) piece of white cardboard.

Procedure. The girls were seated at widely spaced places at long tables, given their materials, and asked to spend some time making a design for the experimenter.[3] The experimenter then demonstrated gluing shapes onto paper, and the children practiced with some scrap materials. None appeared to have any difficulty in working with the materials. Following this, the children were told that they could use the materials in any way they wished to make a design that was "silly." This silliness theme was used as a way to obtain a relatively high baseline level of creativity and to reduce one source of variability—the themes children might employ—that could serve to make the judges' task more difficult. After 18 minutes (at which time virtually everyone had finished), the children were asked to stop.

All judges were told that the designs had been made by children in 18 minutes. Each psychologist judge was individually asked to rank the designs from most to least creative, using his or her own subjective definition of creativity. The art teachers were shown professionally made slides of the 22 designs and were also asked to use their own subjective definitions of creativity in making assessments. They were asked to assign each design to one of five categories immediately after viewing it: (a) very uncreative (b) rather uncreative (c) undecided (d) rather creative or (e) very creative.

The artist judges were recruited to spend 4 hours in an individual session judging the 22 designs on a variety of dimensions. As with the other two groups of judges, the artists were allowed to inspect all 22 designs before making any judgment on any one design and were told that the designs had been made by children in 18 minutes. They evaluated each of the 22 designs on 23 dif-

---

[1] Of course, in the extremes, experience-related variables will be important regardless of the particular task chosen. For this reason, care should be taken in eliminating extremes in ability level on the task in question. For example, if a verbal task was being used, it would be inappropriate to include an age range of 5–15 years in the population for a given study, even if the task had been chosen to minimize the importance of special verbal skills. Likewise, if an artistic task was being used with a population of college students, it would be inappropriate to include students who had been doing studio art for several years along with students who have no involvement in art, even if the task does not depend in an obvious way upon draftsmanship.

[2] It is difficult to provide a clear rule for the number of judges to be used. By the Spearman-Brown reliability calculation (Nunnally, 1967), the mean judge correlation and the number of judges both contribute to the interjudge reliability. Thus, to the extent that the judging is a difficult task and the mean interjudge correlation is low, the number of judges should be increased. However, if the mean correlation is high, good reliabilities can be obtained with fewer judges.

[3] Of the artistic creativity studies presented here, this is the only one in which subjects worked in groups. They were, however, encouraged by the experimenters to "do their own thing" and not to copy or interfere with the other girls. The experimenters were present throughout the session and observed that there was very little imitation of ideas or techniques.

Table 1
Factor Analysis on 23 Dimensions of Artistic Creativity Judgment: Study 1

Dimension	Factor loading: Factor 1: Creativity	Factor loading: Factor 2: Technical Goodness
Creativity cluster		
Creativity	.68	−.23
Novel use of materials	.78	−.21
Novel idea	.55	−.18
Effort evident	.85	.23
Variation in shapes	.72	−.04
Detail	.95	.09
Complexity	.91	−.30
Technical cluster		
Technical goodness	.16	.54
Organization	−.08	.67
Neatness	−.34	.51
Planning	.10	.83
Representationalism	.00	.95
Symmetry	−.34	.48
Expression of meaning	−.01	.92
Aesthetic judgment		
Liking	.22	−.04
Aesthetic appeal	−.04	.14
Would you display it?	.22	.28

ferent dimensions including creativity, technical goodness, and aesthetic appeal. A nonrestrictive definition of each dimension was provided at the top of its rating sheet (e.g., "Creativity: Using your own, subjective definition of creativity, the degree to which the design is creative."). The judges were asked to keep the different dimensions of judgment as separate from one another as possible, and they were instructed to rate the collages relative to one another on each dimension rather than rating them against some absolute standard of art.

The artist-judges were presented with continuous scales for making each judgment. These scales had five equally spaced reference points marked, three of which were labeled "high," "medium," and "low." The judges were told to make each judgment by placing an X anywhere on the scale. They were asked to view the scale as having equal spacing between the five reference points, and they were encouraged to make use of the entire range of the scale. On four of the dimensions (creativity, technical goodness, liking, and silliness), the artist-judges were also asked to assign each design to one of three categories (high, medium, low) and to rank all 22 designs from highest to lowest on the given dimension.

The designs were arranged in a different random order for each artist-judge, and the dimensions were judged in different random orders (e.g., all designs judged on balance, then all judged on creativity, and so on). The various judgment tasks were described to judges beforehand, and the judges were given loose time limits to help them pace themselves through the judgment tasks.

Results

Interjudge reliabilities (Winer, 1971) were calculated for each of the three groups of judges.[4] The reliability of the rankings on creativity provided by the 12 psychologist judges was .73; that of the creativity ratings made by the 21 art teachers was .88. Interjudge reliabilities for the seven artist-judges on the 23 dimensions of judgment were, on the whole, quite good. The reliability of the creativity ratings was .77 and that of the technical goodness ratings was .72. In addition, 16 of the 23 dimensions had reliabilities greater than .80; the median reliability was .77. There was little systematic difference between the reliabilities of the scale ratings, the rankings, and the category groupings for those four dimensions that were assessed in all three ways. Thus, in subsequent studies, only the scale ratings were used.

The agreement between different groups of judges on their creativity assessments was fairly good. The correlation between the psychologist-judges' mean creativity ranking for each design and the artist-judges' mean creativity ranking was .44, $p < .05$. The correlation between the art teachers and the artist-judges was even higher: .65, $p < .01$.

Several of the dimensions assessed by the artist-judges did, indeed, correlate with their judgments of creativity. In addition, however, it appears that the creativity judgments were psychologically separate for the judges from their assessments of technical competence and aesthetic appeal. A factor analysis done on the mean ratings of all 23 dimensions of judgment for each design (varimax rotation) revealed two major factors that were essentially orthogonal to one another.

[4] The technique of calculating interjudge reliabilities presented by Winer (1971) involves an analysis of between-design and within-design variance in ratings, and the result depends on the number of judges used. The Spearman-Brown calculation of reliabilities (Nunnally, 1967), which also depends on the number of judges, yields results for these analyses that are virtually identical to those obtained by the analysis of variance technique. Because the Spearman-Brown calculations are considerably simpler, they were used in subsequent studies.

These appear to be a Creativity factor and a Technical Goodness factor. Many of the 23 dimensions clustered rather neatly about these two factors; loadings on the factors are presented in Table 1. It is important to note that liking for and aesthetic appeal of the collages loaded low on both of these main factors, as did the rated "silliness" of the design.

A number of objective measures were made on the collages by two independent raters: number of pieces used, number of colors used, number of shape categories used (circles, squares, crescents, and so on), number of pieces altered in some way (ripped, folded, and so on), and number of pieces that overlapped other pieces, in addition to several other objective features of the collages. Interrater agreement on these measures was nearly 100%. Many of the objective features did indeed correlate significantly with the artist-judges' ratings of creativity: number of colors, $r = .48$, $p < .05$; number of pieces, $r = .64$, $p < .01$; shape categories, $r = .52$, $p < .05$; pieces altered, $r = .37$, $p < .05$; and pieces overlapping, $r = .62$, $p < .01$.

Importantly, age of the child did not correlate significantly with any of the groups of judges' assessments of creativity, although it did correlate with the rated technical goodness of the collages, $r = .46$, $p < .05$.

Studies 2 and 3

Method

These studies also employed children as subjects. Study 2 included 55 boys and 56 girls drawn from Grades 2–6 at a parochial school in eastern Massachusetts. Study 3 included 24 boys and 23 girls enrolled in Grades 1 and 2 at a nontraditional "open" school in eastern Massachusetts. In Study 2, the judges were six undergraduate studio art majors from Brandeis University, each with at least 4 years of studio art experience. In Study 3, the judges were seven artists from Brandeis University[5] and seven nonartists (psychology graduate students, undergraduates, and elementary-school teachers from schools other than that in which the study was conducted). Using the same materials as in Study 1, each child in these studies worked individually on making a collage for about 15 minutes. Following the standard procedures, each judge rated each collage on creativity and technical goodness.

Results

In Study 2, the interjudge reliabilities were .77 for creativity and .72 for technical goodness. Again, the correlation between subject age and rated creativity of the collage was nonsignificant, $r = .12$, whereas the correlation between subject age and rated technical goodness was significant, $r = .28$, $p < .01$. There were no significant differences between boys and girls in rated creativity, $t(109) = .68$.

In Study 3, the interjudge reliability for creativity judgments was .81 for the artists, .83 for the nonartists, and .89 when all 14 judges were considered together. The reliability for technical goodness judgments was .72 for the artists, .80 for the nonartists, and .83 for all 14 judges. For creativity judgments, the correlation between artists and nonartists was .69, $p < .001$. Over all judges, the correlation between creativity and technical goodness judgments was .77.

Studies 4, 5, and 6[6]

Method

The subjects in each of these studies were college undergraduates: 95 women at Stanford University in Study 4, 10 men and 10 women at Stanford University in Study 5, and 40 women at Brandeis University in Study 6. Using the standard materials, each subject was given 15 minutes to make a collage that "conveyed a feeling of silliness." In addition, in Study 5, half of the male and half of the female subjects were asked to "be as creative as possible."

The judges were 15 artists with at least 5 years of studio art experience in Study 4, 14 nonartist undergraduate and graduate students in Study 5, and 10 students working on honors projects in studio art in Study 6. All judges rated the collages from their study individually, according to the standard procedure. In Study 4, the judges rated each collage on each of 16 different artistic dimensions, which had been formed by eliminating from the 23 dimensions used in Study 1 those that had low interjudge reliabilities or had not clustered with the creativity or technical goodness judgments.

Results

The interjudge reliability of the creativity judgments in Study 4 was .79. In general, the reliabilities of all the subjective judgments in this study were quite high; 12 of the 16 dimension reliabilities were above .80, and the median reliability was .84. Only one dimension of judgment, balance, failed to reach an

---

[5] Some of these judges had also participated in Study 2.

[6] A more complete report of Study 4 can be found in Amabile (1979).

acceptable reliability level. As in Study 1, a factor analysis was performed on the dimensions of judgment. With three exceptions, the 16 dimensions clustered almost exactly as they had in the earlier analysis. There were two nearly orthogonal factors: one composed of creativity, novel material use, novel idea, effort evident, variation of shapes, detail, and complexity; and one composed of neatness, organization, planning evident, balance, and expression of meaning. The asymmetry dimension (the negative of the symmetry judgments) loaded sufficiently low on the creativity factor that it did not cluster with the others. In addition, representationalism loaded nearly zero on both factors, and the single dimension of technical goodness loaded high not only on the technical factor but also on the creativity factor.

In Study 5, the judges rated the collages only on creativity; the reliability of these judgments was .93. Instructions to "be creative" had no impact on the rated creativity of the collages ($F < 1$), but there was a trend toward a sex difference; females made collages that were rated higher in creativity than those of the males, $F(1, 16) = 4.39$, $p < .10$. There were no significant differences between the ratings assigned by female judges and those assigned by male judges, $t(19) = 1.47$.

Both creativity and technical goodness were rated in Study 6, and the reliabilities were .93 and .91, respectively. The correlation between the two dimensions was .70.

Discussion of Collage Studies

This set of studies on the development of a reliable subjective method for assessing artistic creativity has demonstrated several strengths of the methodology. Of primary importance, these studies have shown that it is possible to obtain high levels of agreement in subjective judgments of creativity, even when the judges are working independently and have not been trained to agree in any way. Thus, according to the consensual definition of creativity presented earlier, this can be considered a valid measure of creativity in these products.

It appears, however, that the levels of interjudge agreement for this task may depend to some degree on the magnitude of effort required of judges. For artists, the correlation between the total number of dimensions judged and the mean interjudge correlation of the creativity judgments is −.43. For all judge groups, this correlation is −.24. Likewise, for artists, the correlation between number of collages judged at a given time and mean interjudge correlation of the creativity judgments is −.33. For all judge groups, this correlation is −.14. This is perhaps not surprising, because in Study 4, where each artist spent approximately 4 hours rating 95 collages on each of 16 dimensions, the mean interjudge correlation was only .21; by contrast, in Study 6, where each artist-judge spent only about ½ hour rating 40 collages on 2 dimensions, the mean interjudge correlation was .57. Possible factors of judge fatigue and difficulty in maintaining consistent criteria throughout the judging task should be taken into account by increasing the number of judges used under such circumstances.

It is interesting that level of expertise of the judges appears not to matter as much as might have been expected for this task. Over all six studies, there is no clear superiority of artists over nonartists in interjudge reliability. The mean interjudge correlation is .35 for the former group and .33 for the latter group. Moreover, nonartists and artists do not appear to be subjectively defining creativity in very different ways. In those studies where both types of judges were used (Studies 1 and 3), the degree of agreement between the groups is quite high (.44 and .65 in Study 1, .69 in Study 3). This, of course, raises the question of who is to be considered an expert for the purposes of the present methodology and, thus, who is to be considered an appropriate judge. It appears that the only requirement is a familiarity with the domain of endeavor in which the product was made. Apparently, all the groups of judges used in these collage studies shared some requisite minimal level of familiarity with collages. Indeed, it is probably true that for some domains—judging cartoon captions, for example—any "ordinary" individuals with an average level of exposure to the written media would be appropriate judges. On the other hand, there are other areas—for example, medical research or atonal music—in

which, to be an appropriate judge, an individual must have received special training in the field. Otherwise, the judges' familiarity with the domain would be in doubt and their level of agreement would almost certainly suffer.

These studies have also demonstrated that subjective judgments of the creativity of artworks can be separated from judgments of their aestheic appeal and technical goodness. In Study 1, aesthetic judgments loaded very low on both the Creativity and the Technical Goodness factors in the factor analysis. In both Study 1 and Study 4 (the two studies in which enough dimensions were judged to warrant a factor analysis), there was a clear separation between the Technical Goodness and Creativity factors. It must be noted, however, that the single dimension labeled Technical Goodness does not seem to consistently capture the technical component of subjective judgments. Although the correlation between the single Creativity and Technical Goodness dimensions was only .13 and .26 in Studies 1 and 2, respectively, it was .77, .68, and .70 in Studies 3, 4, and 6, respectively. Because the factor analyses did consistently yield a good separation between the Creativity and Technical Goodness *factors*, it may be that a scale labeled something other than Technical Goodness would better capture the distinct set of features that includes neatness, organization, planning, and so on, and that ratings on such a scale would be more consistently distinct from ratings on the Creativity scale.

Although they were independent of the technical and aesthetic judgments, the creativity ratings were correlated with several other subjective judgments and a number of objective features of the collages as well. A clearly delineated set of artistic dimensions clustered closely with creativity in a factor analysis of subjective judgments, and in Study 1, several physical features of the collages correlated with assessments of creativity.

It is of particular significance for a social psychology of creativity that it is possible, as demonstrated in these studies, to find a creativity assessment task that does not depend heavily on the special skills that some individuals may have developed more highly than others. The collage task was quite manageable for both adults and children and, among the children, there was no apparent relationship between age and creative performance. Thus, this type of task could be used in experimental studies of social and environmental influences on creativity.

Application to Verbal Creativity

Study 7

Method

Subjects and judges. The subjects were 48 female students enrolled in an introductory psychology class at Brandeis University. Two groups of judges assessed the finished products, which were five-line stanzas. Group 1 was composed of seven males and three females recruited from the English department of Brandeis University; all had had at least three years' experience studying poetry at an advanced level and writing poetry themselves. Several were published poets. Group 2 was composed of five published poets and five nonpoets who lived in Cambridge, Massachusetts. The five nonpoets were of approximately the same age and background as the poets.

Procedure. Each subject was given 20 minutes to work individually on writing a cinquain.[7] A cinquain is a simplified form of unrhymed poetry consisting of 5 lines: Line 1 is a single noun; Line 2 consists of two adjectives describing the noun; Line 3 consists of three verb forms relating to the noun; Line 4 contains any number of words (a phrase or sentence about the noun); Line 5 repeats the noun of Line 1. After the initial instructions, subjects were presented with two examples of cinquains. All subjects were provided with the first line of the poem they were to write—the word *joy*—in an effort to reduce variability and to make the judging task somewhat easier.

The judges in Group 1 participated together in one session. Of course, since the judging task involved reading, each judge was able to work completely independently. Indeed, the judges were not allowed to discuss the poems or the judging task at any time before the session was completed. They were each given copies of the instructions that subjects had received and were told that the female undergraduate subjects had had 20 minutes in which to write their poems. Each judge had copies of the 38 poems and was allowed to read through all of them prior to making any judgments. These judges all rated each poem on creativity, using their "own subjective definition of creativity." The judges in Group 2, also participating in one group session, rated each of 24 poems on 14 different dimensions, including creativity, use of the poetic form, and liking.[8] Scales similar to those

[7] Pretesting indicated that 20 minutes was ample time for subjects to finish.

[8] Half of the original 48 poems were randomly chosen for assessment by Group 2. This was done because the task of rereading each poem for rating on each of the

Table 2
Factor Analysis on 14 Dimensions of Judgment for Group 2: Study 7

Dimension	Factor loading		
	Factor 1: Creativity	Factor 2: Style	Factor 3: Technical
	Creativity cluster		
Creativity	.91	.30	.21
Novelty of word choice	.91	.25	.24
Originality of idea	.88	.36	.06
Sophistication	.90	.31	.13
Rhythm	.78	.36	.10
	Style cluster		
Clarity	.23	.89	.27
Appropriateness	.47	.82	.20
Consistency	.44	.81	.16
	Technical cluster		
Grammar	.19	.17	.90
Use of form	.14	.22	.91

used in the collage judging were employed. Indeed, the judging procedure for both Group 1 and Group 2 was virtually identical to that used in the collage studies.

Results

The interjudge reliability of almost all subjective judgments was quite high. The reliability of creativity judgments was .87 for Group 1 and .90 for Group 2. For the 14 dimensions rated by Group 2, 13 of the reliabilities were above .70; the median reliability was .82. In addition, in Group 2, the correlation between the creativity judgments of poets and nonpoets was quite high: .80. A factor analysis (varimax rotation) of the 14 dimensions rated by Group 2 did not produce as clear a separation as was obtained with judgments on the collages. All of the dimensions loaded positively on the two main factors, and some loaded rather high on both. However, there did seem to be some meaningful clusters of dimensions (see Table 2). Liking judgments did not cluster closely with the five creativity dimensions because the former loaded fairly high on both the Creativity and the Style factors. However, the correlation between judges' creativity ratings and their liking ratings was .94.

14 dimensions was quite time consuming. This procedure was preferable to one in which each poem was rated on all 14 dimensions before the next poem was rated on any dimension, however, because the former procedure made it more likely that each judge would apply his or her subjective criteria for each particular dimension more consistently.

Study 8

Method

Subjects in this study were 40 undergraduate females recruited from an introductory psychology class at Brandeis University. They wrote cinquains under the same procedure as had been used for Study 7. The judges were six poets living in the Cambridge, Massachusetts area.[9] These judges worked individually on the assessment tasks in their homes; they were mailed copies of the 40 poems and were asked to follow the same judging procedures as had been followed in the group session for Study 7. In addition, they were given written instructions identical to those used in the previous study. Each judge rated each of the 40 poems on creativity, clarity of expression, and use of poetic form.

Results

Although somewhat lower than the interjudge reliabilities obtained in Study 7, two of the three reliabilities here did meet acceptable levels. The reliability of creativity judgments was .77; poetic form judgments,

[9] Five of these six judges had served in Study 7.

.91; and clarity judgments, .62. Creativity judgments correlated −.16 with poetic form judgments and .38 with clarity judgments; both correlations are nonsignificant.

Discussion of Poetry Studies

This preliminary attempt to extend the proposed assessment technique to the verbal domain was successful on many counts. First, and most important, subjective judgments of creativity of the poems were shown to have high interjudge reliability. In addition, although the separation was not quite as clear as it had been with the artistic judgments in Studies 1 and 4, there was a separation between the Creativity factor and other factors rated in a factor analysis of the subjective judgments in Study 7. As noted earlier, although ratings of liking for the poem did not cluster closely with the creativity dimensions on that factor analysis, the single liking dimension was highly correlated with the single creativity dimension. Thus, it appears that although it may be possible for some types of products to obtain creativity judgments that are clearly uncontaminated by assessments of liking, creativity judgments for other types of products may be more tightly bound up with assessments of aesthetic appeal.

As with the collages, it appears that a very high level of judge expertise is not required to ensure a high level of reliability in ratings of the cinquains. The level of agreement in creativity ratings between nonpoets and published poets in Group 2 of Study 7 was .80. This high level of agreement might have arisen because the cinquain is so simple and because most educated individuals in our culture are familiar with simple poetic forms. As noted earlier, higher levels of expertise would probably be required to ensure reliability of judgments of very complex products or products that would not be generally familiar.

It is also important to note that, like the collage-making task, the cinquain task appears not to depend heavily on special skills. None of the subjects in either study expressed any difficulty in writing the poem or in completing it easily in the time allotted. In addition, these studies demonstrated that this assessment technique can be adapted for very different kinds of tasks. The collage making is completely nonverbal, involving the manipulation of paper and glue. In contrast, the task of writing a cinquain requires the composition of three original lines. It seems probable, therefore, that this methodology could be successfully employed for a variety of tasks in a wide range of domains and that it could be generally useful in experimental studies on the social psychology of creativity.

General Discussion

Contrast to Previous Methods

In its basic approach, the methodology described here is the conceptual reverse of that adopted by traditional objective creativity tests. In those tests, component tasks and subtasks are scored to yield a global assessment of an individual's creativity, an assessment that is ultimately based on the subjective judgment of the psychometricians who devised the subtasks or the raters who score them. Instead, the present technique begins with a global, explicitly subjective assessment of creativity. This global judgment is then clearly demonstrated to be a reliable one. Once this is done, the judgment of creativity can tentatively be broken down into component parts; it can then be examined to determine which other subjective judgments and, perhaps, which objective features of the product predict this judgment of creativity. As the studies presented here demonstrate, some progress toward this goal has already been made.

Moreover, the subjective assessment technique proposed here appears to have ecological validity as a measure of creativity. Using materials that allow for considerable flexibility in response, subjects actually create something—a collage or a poem, for example—that "real world" creators might make. Then, using an approach that has clear precedents in social-psychological research (e.g., Thurstone & Chave, 1929; Walster et al. 1966), reliable subjective judgments of the products are obtained from appropriate observers. Thus, not only does the task itself mimic real-world performance but the assessment technique mimics real-world evaluations of creative work.

The present methodology can be considered a refinement of the subjective assessment techniques used in previous social-psychological and personality studies (e.g., Domino, 1974; Helson & Crutchfield, 1970; Kruglanski et al., 1971; MacKinnon, 1962; Sobel & Rothenberg, 1980). In comparison with those techniques, the present methodology can be more generally useful for social-psychological studies for two reasons. First, it calls for tasks that are structured to be relatively independent of skills such as writing or drawing ability. Second, in contrast to many subjective assessment studies in which judges make global ratings of a *person's* creativity, in this technique judges make ratings of specific *products*. In addition, the reliability of creativity judgments across several groups of appropriate observers has been firmly established in the present method, in the absence of any attempts to train judges to agree. Finally, and perhaps most importantly, the existence of a unique subjective construct called *creativity* has been demonstrated. Although judges were not provided with a definition of creativity, but were instead asked to use their own subjective definitions, they consistently and reliably identified a quality in both types of products that was distinct from technical execution. Moreover, for artworks, it was distinct from aesthetic appeal as well.

It is important that this assessment methodology has been used in several studies investigating substantive questions about creativity. In particular, the methodology has been successfully applied in studies of task-motivational influences on creativity. In one such investigation (Amabile, 1979), subjects making collages either did or did not expect that their work would be evaluated by experts. Those expecting external evaluation not only expressed a lower level of intrinsic interest in the activity but also produced collages that were significantly lower on judged creativity than those produced by nonevaluation subjects. Another study (Goldfarb, Note 1) replicated this effect on the creativity of cinquains.[10] Thus, the construct validity of this subjective method of creativity assessment has been reinforced; the method provides measures of a construct that behaves as creativity was predicted to behave on the basis of theoretical derivations.

Limitations of the Methodology

Despite its advantages, the present technique of creativity assessment is not proposed as an ultimate and universally useful methodology. Indeed, it has some specific limitations. First, if practical considerations are primary, this method is decidedly impractical in the short run. Choosing an appropriate task and an appropriate judge population, having several individuals judge the products (sometimes on several dimensions), and doing statistical analyses are all time consuming.

In addition, to the extent that the task chosen is one that does not depend heavily on particular abilities and experience-related skills, the approach is probably not a useful one for identifying a wide range of enduring individual differences in creativity. However, because tasks can be chosen with this end in mind (i.e., with the inclusion of an appropriate range of skills to be tapped), with modification this method might prove to be a reasonable one even for such individual-difference studies.

Also, it may be difficult to apply this assessment technique to products that are at the frontiers of a particular domain of endeavor. Consider, for example, revolutionary theories in science or revolutionary works of art. It would be difficult to apply this method to assess the creativity of these products because it is precisely their revolutionary nature that makes it difficult for people, even supposed experts in their fields, to agree on the level of creativity evident. In fact, this problem could be considered within the context of the familiarity criterion proposed earlier. These products are so different that no one is sufficiently familiar with the domain (perhaps because the products create their own new domain) to serve as an appropriate judge.

Finally, a related caveat is that the reliability—and hence the validity—of the judgments obtained by this method are necessarily limited by historical time and place. It is doubtful that a group of Italian Renais-

[10] In addition, although their subjective assessment methodology was considerably less complex than that outlined here, Kruglanski et al. (1971) found negative effects of expected reward on the rated creativity of stories and titles.

sance painters would agree to a great degree with a group of contemporary American artists in their creativity judgments of any given group of artworks from any era. Clearly, the shared subjective criteria of creativity in any domain of endeavor do change over time and do differ across cultures. This need not necessarily be considered a limitation, however. Many previous theorists have realized that the judgment of creativity is historically and culturally bound,[11] and it seems unreasonable to expect that universal and enduring criteria—even subjective criteria—could ever be agreed upon.

Future Directions

Although the preliminary studies reported earlier lay the groundwork for the assessment methodology proposed here, further refinement and extension of the technique are required in a number of areas. First, attempts should be made to extend the subjective assessment methodology to other domains of endeavor and to other types of tasks within the artistic and verbal domains already explored. Second, a finer identification of objective and subjective correlates of creativity judgments may be possible. Third, the limits of interjudge reliability should be explored, especially where the products being judged are in relatively new domains or represent truly pioneering work. Fourth, since the results presented earlier suggested that certain characteristics of the judging task—such as the number of products or dimensions to be judged—can influence interjudge agreement, it is important to determine what other features of the judging task might also influence reliability. Fifth, more work is needed on the identification of appropriate judges for particular types of products and on the influence of judge characteristics on interjudge reliability. Sixth, because the studies presented here showed reasonable but not perfect separation between creativity and other dimensions of subjective judgment, it is important to determine under what circumstances and for what types of products and domains of endeavor creativity judgments will and will not be clearly separated from assessments of the technical and aesthetic aspects of products. Finally, attempts should be made to determine whether this subjective assessment technique could be usefully adapted not only for the social psychology of creativity but also for the identification of individual differences in creativity.

The arguments presented here on the merits of various creativity-assessment techniques for different types of psychological inquiries raises a more general point about the aims and methods of personality and social psychology. Most simply stated, it is the purpose of personality psychology to identify stable characteristics and abilities and to study the ways in which they are related to other characteristics and abilities. Thus, personality researchers generally look for consistency of behavior over time and across situations and for evidence of a given trait in a variety of behavioral domains. For these purposes, measures that are particularly sensitive to stable baseline individual differences are most useful. On the other hand, most simply stated, it is the purpose of social psychology to investigate the impact of social and environmental factors on most people or on the "average" person. Thus, social psychologists are less concerned with within-group variability (and in fact seek to reduce it where possible) and more interested in situationally induced between-group differences in behavior. Such differences, although they may be large and important ones, are usually of relatively brief duration when compared to the enduring individual differences that personality researchers seek to identify. For their purposes, then, social psychologists require measures that are not particularly sensitive to baseline individual differences but do allow room for variability in response. This difference between appropriate methods applies not only to creativity assessment but also to measurement in a wide range of other behavioral domains that are of interest to both social and personality psychology.

Rather than proposing a new and easily applicable creativity test, the consensual assessment technique instead presents a somewhat time-consuming general methodology that can produce clear and reliable subjective

[11] For example, Stein (1975, p. 253) defines creativity as "a process that results in a novel product or idea which is accepted as useful, tenable, or satisfying by a significant group of others at some point in time."

judgments of creativity. In the long run, though, more than any single creativity test, this methodological approach may facilitate the inclusion of creativity in a comprehensive theory of social-psychological influences on performance and the inclusion of social psychological influences in a comprehensive theory of creativity.

Reference Note

1. Goldfarb, P. *The effects of evaluation expectation and coaction on verbal creativity.* Unpublished manuscript, Brandeis University, 1978. (Available from T. Amabile, Department of Psychology, Brandeis University, Waltham, Massachusetts 02254).

References

Amabile, T. M. Effects of external evaluation on artistic creativity. *Journal of Personality and Social Psychology*, 1979, *37*, 221–233.

Amabile, T. M. The social psychology of creativity: A componential conceptualization. *Journal of Personality and Social Psychology*, in press.

Barron, F. The disposition toward originality. *Journal of Abnormal and Social Psychology*, 1955, *51*, 478–485.

Bruner, J. The conditions of creativity. In H. Gruber, G. Terrell, & M. Wertheimer (Eds.), *Contemporary approaches to creative thinking.* New York: Atherton Press, 1962.

Carlsmith, J., Ellsworth, P., & Aronson, E. *Methods of research in social psychology.* Reading, Mass.: Addison-Wesley, 1976.

Condry, J. Enemies of exploration: Self-initiated versus other-initiated learning. *Journal of Personality and Social Psychology*, 1977, *35*, 459–477.

Deci, E. Effects of externally mediated rewards on intrinsic motivation. *Journal of Personality and Social Psychology*, 1971, *18*, 105–115.

Diener, C., & Dweck, C. An analysis of learned helplessness: Continuous changes in performance, strategy, and achievement cognitions following failure. *Journal of Personality and Social Psychology*, 1978, *36*, 451–462.

Domino, G. Assessment of cinematographic creativity. *Journal of Personality and Social Psychology*, 1974, *30*, 150–154.

Eisner, E. A typology of creative behavior in the visual arts. *American Educational Research Journal*, 1965, *2*, 125–136.

Feldman, D. *Beyond universals in cognitive development.* Norwood, N.J.: Ablex, 1980.

Gakhar, S., & Luthra, S. The test–retest reliability of Torrance Tests of Verbal Creative Thinking in a sample of ninth and tenth grade Indian children. *Journal of the Indian Academy of Applied Psychology*, 1973, *10*, 48–52.

Getzels, J., & Jackson, P. *Creativity and intelligence: Exploration with gifted students.* New York: Wiley, 1962.

Ghiselin, B. Ultimate criteria for two levels of creativity. In C. Taylor & F. Barron (Eds.), *Scientific creativity: Its recognition and development.* New York: Wiley 1963.

Gough, H. A creative personality scale for the adjective check list. *Journal of Personality and Social Psychology*, 1979, *37*, 1398–1407.

Guilford, J. P. The structure of intellect. *Psychological Bulletin*, 1956, *53*, 267–293.

Guilford, J. P. A revised structure of intellect. *Report of the Psychological Laboratory*, University of Southern California, No. 19, 1957. (a)

Guilford, J. P. Creative abilities in the arts. *Psychological Review*, 1957, *64*, 110–118. (b)

Guilford, J. P. Traits of creativity. In H. Anderson (Ed.), *Creativity and its cultivation.* New York: Harper, 1959.

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

Helson, R., & Crutchfield, R. Mathematicians: The creative researcher and the average PhD. *Journal of Consulting and Clinical Psychology*, 1970, *34*, 250–257.

Holland, J. Test reviews. *Journal of Counseling Psychology*, 1968, *15*, 297–298.

Kruglanski, A., Friedman, E., & Zeevi, G. The effects of extrinsic incentives on some qualitative aspects of task performance. *Journal of Personality*, 1971, *39*, 606–617.

Lepper, M., Greene, D., & Nisbett, R. Undermining children's intrinsic interest with extrinsic reward: A test of the "overjustification" hypothesis. *Journal of Personality and Social Psychology*, 1973, *28*, 129–137.

Lott, B. Behavioral concordance with sex role ideology related to play areas, creativity, and parental sex typing of children. *Journal of Personality and Social Psychology*, 1978, *36*, 1087–1100.

MacKinnon, D. The nature and nurture of creative talent. *American Psychologist*, 1962, *17*, 484–495.

McGraw, K., & McCullers, J. Evidence of a detrimental effect of extrinsic incentives on breaking a mental set. *Journal of Experimental Social Psychology*, 1979, *15*, 285–294.

Newell, A., Shaw, J., & Simon, H. The processes of creative thinking. In H. Gruber, G. Terrell, & M. Wertheimer (Eds.), *Contemporary approaches to creative thinking.* New York: Atherton Press, 1962.

Nicholls, J. Creativity in the person who will never produce anything original and useful: The concept of creativity as a normally distributed trait. *American Psychologist*, 1972, *27*, 717–727.

Nunnally, J. *Psychometric theory.* New York: McGraw-Hill, 1967.

Seta, J., & Hassan, R. Awareness of prior success or failure: A critical factor in task performance. *Journal of Personality and Social Psychology*, 1980, *39*, 70–76.

Simonton, D. K. Sociocultural context of individual creativity: A transhistorical time-series analysis. *Journal of Personality and Social Psychology*, 1975, *32*, 1119–1133.

Simonton, D. K. Creative productivity, age, and stress: A biographical time-series analysis of 10 classical composers. *Journal of Personality and Social Psychology*, 1977, *35*, 791–804. (a)

Simonton, D. K. Eminence, creativity, and geographic marginality: A recursive structural equation model. *Journal of Personality and Social Psychology*, 1977, *35*, 805–816. (b)

Simonton, D. K. Multiple discovery and invention: Zeitgeist, genius, or chance? *Journal of Personality and Social Psychology*, 1979, *37*, 1603–1616.

Sobel, R., & Rothenberg, A. Artistic creation as stimulated by superimposed versus separated visual images. *Journal of Personality and Social Psychology*, 1980, *39*, 953–960.

Stein, M. *Stimulating creativity* (Vol. 1). New York: Academic Press, 1974.

Stein, M. *Stimulating creativity* (Vol. 2). New York: Academic Press, 1975.

Thurstone, L., & Chave, E. *The measurement of attitude.* Chicago: University of Illinois Press, 1929.

Torrance, E. P. *Guiding creative talent.* Englewood Cliffs, N.J.: Prentice-Hall, 1962.

Torrance, E. P. Career patterns and peak creative achievements of creative high school students twelve years later. *Gifted Child Quarterly*, 1972, *16*, 75–88. (a)

Torrance, E. P. Predictive validity of the Torrance Tests of Creative Thinking. *Journal of Creative Behavior*, 1972, *6*, 236–252. (b)

Wallach, M., & Kogan, N. *Modes of thinking in young children.* New York: Holt, Rinehart & Winston, 1965.

Walster, E., Aronson, V., Abrahams, D., & Rottman, L. Importance of physical attractiveness in dating behavior. *Journal of Personality and Social Psychology*, 1966, *4*, 508–516.

Wimperis, B., & Farr, J. The effects of task content and reward contingency upon task performance and satisfaction. *Journal of Applied Social Psychology*, 1979, *9*, 229–249.

Winer, B. *Statistical principles in experimental design.* New York: McGraw-Hill, 1971.

Received August 28, 1981 ■

Instructions to Authors

APA policy prohibits an author from simultaneously submitting the same manuscript to two or more journals. Authors should prepare manuscripts according to the *Publication Manual of the American Psychological Association* (2nd edition) and the "Guidelines for Nonsexist Language in APA Journals" (Publication Manual Change Sheet 2, *American Psychologist*, June 1977, pp. 487–494). Manuscripts must include an abstract of 100–175 words. Instructions on tables, figures, references, metrics, and typing (all copy must be double-spaced) appear in the Manual. Articles not prepared according to the Manual will not be reviewed. For further information on content, authors should refer to the editorial in the March 1979 issue of this journal (Vol. 37, No. 3, pp. 468–469).

The reference citation for any article in any JPSP section follows APA's standard reference style for journal articles; that is, authors, article title, journal title, year of publication, volume number, and page numbers. (See Example 1, p. 122 of the *Publication Manual*, 2nd ed.) The citation does *not* include the section title.

Manuscripts should be submitted in quadruplicate (an original and 3 copies) to the appropriate section editor. All copies should be clear, readable, and on paper of good quality. Editors' addresses appear on the inside front cover of the journal. Authors should keep a copy of the manuscript to guard against loss.

Blind review is optional and must be specifically requested when a manuscript is first submitted. In those cases, authors' names and affiliations must appear only on a separate title page, included with each copy of the manuscript. Footnotes containing the identity of authors or their affiliations must be on a separate sheet.

Section editors reserve the right to redirect papers among themselves as appropriate unless an author specifically requests otherwise. Rejection by one section editor is considered rejection by all, therefore a manuscript rejected by one section editor should not be submitted to another.