Measurement of Creativity: Review and Critique.

widely used measures of creativity are reviewed and classified into ten categories: tests of divergent thinking; attitude and interest inventories; personality inventories; biographical inventories; teacher nominations; peer nominations; supervisor ratings; judgments of products; nomination of eminent persons; and self-reported creative activities and achievement. Judgments by peers, supervisors, or teachers are hampered by low interjudge reliability, halo effects, and low discriminant validity. The latter problem is due to an inability to distinguish creativity from other constructs such as intelligence or achievement; or the inability to discriminate among dimensions of creativity, such as fluency, flexibility and inventiveness. Tests and inventories are best described as correlates of real life creative behavior: they explain something about the behavior, but are not acceptable for selection. The most serious measurement problem is a lack of convergent validity among all these methods—correlations between methods are low. A simple and straightforward inventory of creative achievement appears to be the best measure because it is direct, observable, and validated in research as the best predictor of future creative behavior. (Author/JCP)
Measurement of Creativity:
Review and Critique

Dennis Hocevar
University of Southern California

Review presented at the 1979 conference of the Rocky Mountain Psychological Association. Copies of this review may be obtained from Dennis Hocevar, 700 E. 60th St., Los Angeles, CA 90034.
Measurement of Creativity

Abstract

Presently available criteria of creativity are reviewed and classified into ten categories: tests of divergent thinking, attitude and interest inventories, personality inventories, biographical inventories, teacher nominations, peer nominations, supervisor ratings, judgments of products, eminence and self-reported creative activities and achievements. These techniques for measuring creativity are then criticized in terms of their reliability, discriminant validity, dimensionality and convergent validity. It is concluded that an inventory of self-reported creative activities and accomplishments is the most defensible technique for selecting creative individuals.
Perhaps no psychological concept has proven to be as difficult to measure as creativity. A wealth of research in the last twenty-five years has included the development of numerous instruments which are used in the identification of creative talent (see Davis, 1971; Kaltsounis, 1971, 1972). The most salient characteristic of these instruments is their diversity. This diversity is indicative of both the complexity of creativity and the multitude of goals, research designs, subjects, settings, etc. that are characteristic of research on the psychology of creativity. Understandably, a number of experts in the area have suggested that the measurement of creativity be given more attention (Dellas & Gaier, 1970; Treffinger & Poggio, 1972; Treffinger, Renzulli & Feldhusan, 1971; Yamamoto, 1965).

Hopefully, some of the confusion regarding creativity can be dispelled by classifying many of the existing criteria. Accordingly, the different approaches to identifying creativity will be categorized into ten groups. These ten categories are meant only to be descriptive, so consequently, they are neither final nor mutually exclusive. Following these descriptions, a general critique of the existing criteria of creativity will be provided.

**Tests of Divergent Thinking**

Describing creativity in terms of divergent thinking is the most widely used approach to studying creativity. On the basis of Guilford's (1956) structure of the intellect model and over a
decade of factor analytic research, Guilford and his colleagues have identified various intellectual abilities. Some of these abilities (e.g., fluency, flexibility, originality, redefinition, elaboration, etc.) have been collectively labelled divergent thinking. A wide variety of tests have been developed by Guilford and his colleagues to measure divergent thinking, such as Alternate Uses (Christensen, Guilford, Merrifield & Wilson, 1960), Plot Titles (Berger & Guilford, 1969) and Consequences (Christensen, Merrifield & Guilford, 1958). Tests of divergent thinking are distinguished from traditional intelligence tests in that they require a multitude of responses, rather than a single correct answer. For example, in the Alternate Uses tests, subjects are asked to think of alternate uses for a variety of common objects (e.g., shoe, pencil, etc.). Similarly, in the Plot Titles test, subjects are asked to generate clever titles to two stories.

Guilford (1968) and others (Torrance, 1974; Wallach & Wing, 1969) have suggested that the more creative individual should possess the types of abilities measured by tests of divergent thinking, and the Guilford tradition has had considerable impact on the study of creativity. A number of investigators have developed assessment devices (e.g., Getzels & Jackson, 1962; Gough, 1975, 1976; Industrial Relations Center, 1959; Lawshe & Harris, 1960; Mednick, S. & Mednick, M., 1967; Torrance, 1974; Wallach & Kogan, 1965) which may be considered modifications of the Guilford
tests of divergent thinking.

Attitude and Interest Inventories

Some investigators have suggested that creativity can be identified in terms of interests and attitudes. This approach is based on the assumption that a creative person will express attitudes and interests favoring creative activities. For example, the Guilford-Zimmerman Interest Inventory (Guilford & Zimmerman, 1963) contains a "creative interests" scale. In this test, subjects are asked to indicate their interest in a wide variety of activities. As reported by Taft and Gilchrist (1970), a creative individual would indicate interest in items like the following:

- Formulate a new theory to replace one that has been accepted for years.
- Invent a new gadget.
- Write words in new combinations to convey emotion rather than meaning.
- Think up plots for novels.

Similarly, on the Holland and Baird (1968) Preconscious Activity Scale, individuals high on originality agree with items like the following:

- I would rather be a research engineer than an industrial engineer.
- I often daydream about unsolved problems.
- I have to learn things in my own way rather than accepting ideas or relationships suggested in textbooks, etc.
If I had the necessary talent, I would enjoy being a sculptor.

I often act without thinking.

I would like to be an inventor.

The Study of Values (Allport, Vernon & Lindzey, 1960), a scale developed to measure basic interests and motives, has been used to assess creativity by the Institute of Personality Assessment and Research and the Center for Research and Development, both at the University of California at Berkeley (Heist, 1968). Researchers at these two institutions believe that a certain pattern of values typically discriminate the creative individual. The pattern consists of very high scores on the AVL's Theoretical and Aesthetic Scales, and very low scores on the AVL's Economic, Religious, and Social Scales (Heist, 1968).

Other attitude and interest inventories which are purported to be useful in studying creativity are the Runner Studies of Associates, Attitudinal Patterns (Rurner, 1954), the Childhood Attitude Inventory for Problem Solving (Covington, 1966), the Creative Behavior Disposition Scale (Taylor, I., Sutton & Haworth, 1974), Creative Attitude Survey (Schaefer & Bridges, 1970), and the Opinion, Attitude, and Interest Survey (Fricke, 1965).

Personality Inventories

Some investigators have characterized creativity as a set of personality factors rather than cognitive traits. Consequently, several well known personality inventories have been scaled to
identify creativity. For example, in at least four studies, the Adjective Check List (Gough & Heilbrun, 1965) has been suggested as a potential measure of creativity (Domino, 1970; Lacey & Erickson, 1974; Smith & Schaefer, 1969; Welsh, 1967). Some representative adjectives common to the first three of these studies are: clever, complicated, cynical, imaginative, original, reflective, and unconventional (Yarnell, 1971).

Additional personality inventories which have been adapted to identify creativity are Heist and Yonge's (1968) Omnibus Personality Inventory (Heist, 1968), Gough's (1957) California Psychological Inventory (Helson, 1965), Barron's (1953) Independence of Judgment Scale (Schmeidler, 1965), and R. Cattell and Eber's (1968) Sixteen Personality Factor Questionnaire (Cattell, R. & Butcher, 1963). Finally, Torrance and Khatena (1970) have designed a personality inventory specifically for identifying creative adolescents. Items on the "What Kind of Person are You?" instrument call for the test taker to select characteristics in a forced-choice format. For example, the creative person will describe himself as curious rather than self confident, a self-starter rather than obedient, intuitive rather than remembers well, and altruistic rather than courteous. Holmes (1976) has also developed a questionnaire measure of the creative personality.

Biographical Inventories

With the assumption that an individual's present behavior is
determined by past experiences, biographical inventories have been used to assess creative talent. A number of studies indicate that "tailor-made" biographical inventories can predict creativity in an industrial setting (Buel, 1965; Buel, Albright & Glennon, 1966; McDermid, 1965; Owens, Schumacher & Clark, 1957).

Two published inventories with more applicability are presently available: the Alpha Biographical Inventory by the Institute for Behavioral Research in Creativity (1968) and Schaefer's (1970) Biographical Inventory: Creativity. The Alpha Biographical Inventory is the result of an extensive research program carried out with NASA Scientists and engineers (Taylor, W., & Ellison, 1964) and, consequently, it is limited to measuring creativity in the scientific areas. It consists of three hundred items which cover a wide variety of areas such as hobbies, interests, childhood activities, self-ratings, experiences, etc., and it can be scored for both creativity and academic performance.

Schaefer's (1970) Biographical Inventory: Creativity is similar in nature but broader in applicability. One hundred and sixty-five questions are grouped into five sections: physical characteristics, family history, educational history, leisure-time activities, and a miscellaneous category. The inventory has several dimensions measuring different fields of creative endeavor. Boys are ranked on a math-science dimension and an art-writing dimension. Girls are ranked on a writing dimension and an art dimension. Several validation studies have been conducted using
Measurement of Creativity


Teacher Nominations

A large portion of the research on creativity takes place in educational settings. Accordingly, teacher ratings are a commonly used criteria of creativity (e.g., Foster, 1971; Haddon & Lytton, 1971; Piers, Daniels & Quackenbush, 1960; Richards, Cline & Needham, 1964; Yamamoto, 1963).

At the elementary school level, Yamamoto's (1963) work is illustrative. Yamamoto first established standards for teachers to use in their ratings. These standards usually involved identifying the most and least creative thinkers in the class. Creative thinking was defined as fluency (lots of ideas), flexibility (many different ideas), inventiveness (inventing and developing ideas), originality (unique ideas), and elaboration (detailed ideas). Nelson (1963) used a different approach to obtain teacher nominations. Nelson compiled a list of personality traits which were cited in the research literature as characteristic of the creative personality. This list was then used as a checklist for teachers to use in making their judgments.

Teacher nominations have also been used at higher educational levels. Drevdahl (1956) asked faculty members in an arts and science department to rate their students on a seven-point scale of creativity. The raters were asked to make two judgments: one based on their own definition of creativity and one based on
Drevdahl's definition. No significant difference between the defined and undefined ratings were found. In an art school setting, Getzels and Csikszentmihalyi (Note 1) directed the art teachers to rate their students on a four-point originality scale where originality was defined as the "ability to originate ideas and to draw on personal resources in preparing assignments (p. 27)."

They also had the teachers rate the artistic potential of the students where artistic potential was defined as the "capacity for growth and development of innate talent in pursuit of positive professional success in chosen field (p. 27)." A composite of these two ratings was used as a measure of creativity. At the secondary school level, Rivlin (1959) had 25 teachers attend a training conference in which a total of 14 creativity rating standards were discussed. On the basis of these standards, teachers were asked to select five intelligent, creative students and five equally able noncreative students from their classes.

Teacher ratings have also served as criteria on the validation of a number of creativity tests such as Mednick's (1967) Remote Associations Test (Karlins, Schuerhoff & Kaplan, 1969; Mednick, M., 1963) and Schaefer's Biographical Inventory: Creativity (Schaefer, 1969; Schaefer & Anastasi, 1968).

Peer Nominations

While recognizing some inherent limitations in using peer nominations as a criteria of creativity, Torrance (1974) has suggested that they may provide some useful data in the study of
creativity. In order to eliminate the ambiguity in peer nominations, Torrance (1962) instructed young children to base their nominations on specific criteria such as ideational fluency, unusual ideas, problem solving ability, inventiveness, and so on. Yamamoto (1964a, b), Foster (1971) and Reid, King and Wickwire (1959) have also used peer nominations as a criteria of creativity and have developed questionnaires which included similar guidelines.

Surprisingly, peer ratings are generally used with children although there are some exceptions. In working with research scientists, Taylor, C., Smith and Ghiselin (1963) obtained six creativity ratings, rankings, and nominations by peers in a research lab. Sprecher (1964) acquired peer descriptions on twelve characteristics taken to identify creative engineers. These characteristics included: skill in reporting results to others, persistence, likes unusual and challenging problems, analytic thinking, is able to reach a decision on his own, fluency of ideas, working energetically, knows subject matter, friendly relations with co-workers, foresight, develops original approaches to problems, and develops valuable and worthwhile ideas.

**Supervisor Ratings**

Particularly in industrial settings, supervisor ratings have proven to be a useful method of choosing creative employees (Andrews, 1962; Meer & Stein, 1955; Taylor, C., 1958). Supervisors have the advantage of being familiar with the work of many indi-
Measurement of Creativity

11. Individuals and of having the expertise to judge the work of these individuals.

Buel (1960) asked research supervisors in a large oil company to anonymously describe the most and least creative research men under their supervision. On the basis of their descriptions, Buel developed a list of creative behaviors which he says can be used as a descriptive checklist for helping supervisors identify creative research personnel. Taylor and his colleagues at the Institute of Personality Assessment and Research have also employed supervisor ratings in their research with scientists in a government research lab (Taylor, C., Smith & Chiselin, 1963). Immediate supervisors were asked to rate scientists on a number of traits including: productivity, drive, mathematical ability, integrity, desire for facts, independence, informative ability, flexibility, persistence, cooperation, and creation. Taylor also asked laboratory chiefs (higher-level supervisors) to rate scientists with the aid of several checklists.

Judgment of Products

It is self evident that creative people should produce creative products. Understandably, identifying creative people in terms of one or more products is an approach that has encompassed virtually all areas of study and a varied sample of populations.

A little known but ambitious example of this approach is a battery of tests developed by Foster (1971). Foster's assessment devices involve a broad range of activities typically associated
Measurement of Creativity

12.

with the secondary school. A brief description of these activities and their corresponding areas follows:

1. Sorting playing cards in sets of six that belong together (mathematics).
2. Creating mathematical equalities with a given set of numbers and symbols (mathematics).
3. Working in a physical education class with and without equipment (physical education).
4. Playing charades with several pieces of equipment such as masks, hats, tools, etc. (drama).
5. Writing a story (literature).
6. Making a model out of one or several materials such as buttons, feathers, wire, glue, corks, cloth, etc. (model-making).
7. Listening to music and painting a picture of what the music brings to mind (music).

Each of these products were then rated using guidelines which were generally based on three criteria: fluency, flexibility and originality.

Most of the studies incorporating judgments of products have not been as broad as Foster’s work. Usually, they have been limited to judgments of a product in a single area such as art (e.g., Brittain & Beittel, 1964; Csikszentmihalyi & Getzels, 1970; Jones, 1964; Wallbrown, 1972), literature (Jones, 1964; Lynch & Kaufman, 1974; Rookey, 1974; Wallen & Stevenson, 1960), and science
In addition, products are sometimes defined less concretely as ideas (Graham, 1965; Ward & Cox, 1974). The judges in product studies vary from experts to nonexperts, and the criteria vary from diverse definitions of creativity to social recognition (i.e., rewards, publication, etc.). Furthermore, subjective judgments are usually made on what products to choose and in what situation they will be obtained. Consequently, the products generally represent very narrow samples of behavior.

**Eminence**

An impractical but valuable approach to the study of creativity is to study eminent people. Perhaps the most noteworthy research of this type is the work done at the Institute of Personality Assessment and Research (Barron, 1969; Hall & Mackinnon, 1969; Helson, 1971; Helson & Crutchfield, 1970; Mackinnon, 1962). The IPAR group initially asked a panel of professors of architecture to nominate the forty most creative architects in the United States. On the basis of these nominations, sixty-four architects were asked to visit Berkeley for a weekend of intensive study at IPAR. Forty of these architects accepted. At IPAR, groups of creative writers and creative mathematicians were also selected by a similar process.

A number of other studies have attempted to identify men of eminence. Cox (1926) used J. Mck. Cattell's (1903) list of the one thousand most eminent individuals in history. Cattell's
Measurement of Creativity

14.

criteria of eminence was space accorded to each one in biographical dictionaries. Ellis (1904) based his definition of eminence on space devoted to individuals in the Dictionary of National Biography, a British anthology of eminent people. Cattell, R. and Drevdahl (1955) selected eminent researchers, teachers, and administrators in the social sciences and physical sciences on the basis of committee selections and membership in a professional society. Roe (1951a) selected a group of physical scientists by asking a committee of scientists to rate a total of 69 men on a three-point scale of excellence. Roe also chose a group of eminent biologists through a similar means (Roe, 1951b). In Ireland, Barron (1969) asked the staff of the Irish Management Institute, a 6000 member management training group, to select individuals who were unquestionably leaders in Irish Economic life.

The use of eminent people in the study of creativity raises the question of whether studying creativity in the highly gifted is tantamount to studying creativity in "normals." Although most researchers have treated creativity as a normally distributed trait, the argument that it is limited to a very small segment of the population is a worthwhile consideration.

Self-Reported Creative Activities and Achievements

Perhaps the most easily defensible way of identifying creative talent is in terms of self-reported creative activities and achievements. Although there is a problem in deciding which activities and achievements should be designated as creative, most of the
lists that have been used in research have a reasonable degree of face validity. Creative activities can take place in a number of fields which are recognized by society as important. In studying talent for the National Merit Scholarship Corporation, Holland and Nichols (1964) describe lists of achievements and activities in science, art, literature, music, etc. The achievements described by Holland are rare, demand commitment, and are generally publicly recognized through prizes, publications, etc. For example, some illustrative items are: placed first, second or third in a regional or state science contest; exhibited or performed a work of art; had poems, stories or articles published in a public newspaper; invented a patentable device; had minor roles or leads in plays produced by a college or university. Research incorporating this type of checklist has been published in a number of studies (Holland & Astin, 1962; Holland & Baird, 1968; Holland & Nichols, 1964; Holland & Richards, 1965; Richards, Holland & Lutz, 1967a,b).

Other lists similar in content and emphasis to the Holland list have been developed by Erikson (1966), Hocevar (1977), Torrance (1969a,b), Skager, Schultz & Klein (1965) and Wallach and Wing (1969). In general, the total creativity score is simply the number of activities checked. Although these lists have been used in educational and psychological research, there are no commercially available checklists.

As stated earlier, these ten categories are meant only to
identify some common techniques for assessing creative talent. Because of the varied nature of research on creativity, there are other tests of creativity which are tailor made for specific situations, and do not easily lend themselves to categorization. A partial listing follows: Starkweather Creativity Tests (Starkweather, 1971); Ideal Pupil Checklist (Torrance, 1975); Mosaic Construction Test (Hall, 1972); Barron-Welsh Art Scale (Barron & Welsh, 1952); Welsh Figure Preference Test (Welsh, 1959); Lundsteen Test of Creative Problem Solving (Rickborn & Lundsteen, 1968); Onomatopoeia and Images (Khatena, 1969); Pennsylvania Assessment of Creative Tendency (Rookey, 1971); Group Inventory for Finding Creative Talent (Rimm & Davis, 1976); Ingenuity Test (Flanagan, 1968); and a series of cognitive measures developed by researchers at Educational Testing Service (Frederiksen, Evans & Ward, 1975). In the next section some general problems and criticisms of the ten major approaches to the measurement of creativity will be identified.

Discussion and Critique

A number of the approaches identified earlier require people (i.e., supervisors, peers, teachers) to make judgments about products, ideas or other people. This technique presents a particular problem since the researcher must decide who the judges should be and what the judges should be looking for. To illustrate, when judgments of art products are made, there is some evidence that experts and nonexperts disagree. (Golann, 1963;
Knapp & Wulff, 1963). Skager, Schultz and Klein (1965) have identified at least three different points of view regarding art judgments. Despite studies like these, the interjudge reliabilities reported in most creativity research is moderately high (e.g., Bartlett & Davis, 1974; Gough, 1976; Karlins, Schuerhoff & Kaplan, 1969; Moss & Duenk, 1967; Skager, Schultz & Klein, 1965; Wallen & Stevenson, 1960; Ward & Cox, 1974). As a potential solution to problems of interjudge reliability, a strategy for the \textit{a priori} selection of homogeneous judges in a product-centered approach to creativity has been developed by Korb and Frankiewicz (1976).

Related to the problem of interjudge reliability is the problem of how the judges are asked to formulate their decisions. Not surprisingly, different researchers ask their judges to consider different criteria. The list of criteria is virtually unending, including fluency, originality and other cognitive characteristics, personality traits, self-expression, enthusiasm, productivity, and expertise. Some investigators use elaborate definitions of creativity; others use no definition at all. Furthermore, there is no guarantee that judges will understand and be guided by the sometimes complex definitions, and there is little research that shows the instructions even make a difference. To the contrary, when Drevdahl (1956) compared defined and undefined conditions with teacher nominations, no significant differences in the judgments of their most creative students were found. Similarly,
Karlins et al. (1969) reported a correlation of .97 between defined and undefined ratings of creativity.

Another issue is the discriminant validity of judgments. It is reasonable to expect that judges are able to distinguish creativity from other constructs such as intelligence, achievement, competence, etc. A review of some appropriate studies indicates otherwise. Holland (1959) had teachers, principals, and guidance counselors rate students on twelve traits, including originality. Originality correlated .72 with speaking skills and .84 with writing skills. The other nine correlations which involved originality ranged from .50 to .65. Wallen and Stevenson (1960) investigated creativity in fifth grade writing and found that the teacher's judgments correlated .57 with IQ, .66 with school grades, and from .66 to .72 with three standardized achievement tests. In a study of research personnel, instructor ratings of creativeness correlated .68, .72, and .75 with grades, an officership rating and a logical reasoning rating, respectively (Mullins, 1964). Finally, in a study involving faculty ratings of creativity in architecture, Karlins et al. (1969) found that ratings of creativity correlated from .79 to .95 with ratings of dependability, adaptiveness, need to know, independence, and productivity.

Regarding art judgments, the overall results are even more striking. Brittain & Beittel (1964) found that creativity scores determined by judgments on three art performances were correlated .89 with judged aesthetic quality. Csikszentmihalyi and Getzels
(1970) found that the originality ratings of a number of drawings correlated .76 with a rating of technical skill and .90 with a rating of overall aesthetic value. Similarly, in the Getzels and Csikszentmihalyi (Note 1) study, mean ratings on originality correlated with mean ratings on artistic potential .72 for males and .77 for females. Finally, Rossman and Gollob (1975) report that among art students, peer ratings of creativity correlated .84 with peer ratings of intelligence. Likewise, faculty ratings of creativity correlated .59 with faculty ratings of intelligence.

Since judges have trouble discriminating creativity from other attributes, it follows that they will have even more trouble discriminating various dimensions of creativity. In the study cited earlier, Foster (1971) reported that the correlations between fluency, flexibility, and originality ratings were about .80 in physical education, painting, model-making, writing and drama. On the same line, Yamamoto (1964b) found that when peer nominations were used to measure fluency, flexibility and inventive level, their intercorrelations ranged from .62 to .82.

Notably, all of these intercorrelations would be even higher if the data were free of measurement error (i.e., unreliability in the judgments). It is obvious that a considerable "halo" effect is present when creativity is based on subjective judgments. Judges seem to only be able to establish some overall opinion which influences all of their judgments. Because of this, they fail to discriminate creativity. The evidence indicates that
asking judges to choose the most intelligent people or their favorite product is probably equivalent to obtaining judgments of creativity. There is one exception to this general finding. Rossman and Gollob (1975) have demonstrated that subjects can make distinct judgments regarding creativity and intelligence when the judgments are based on diverse information. Specifically Rossman and Gollob gave their raters abilities information, personality information and biographical information. It is important to note that the raters in the Rossman and Gollob (1975) study did not know the individuals that they were rating, thus, making objective judgments more likely and the "halo" effect less likely.

Another disturbing characteristic of most of the techniques for measuring creativity is the almost total neglect of the dimensionality of the attribute. In most instances, creativity is thought of as a unitary trait—there is little differentiation among creativity in the arts, sciences, literature, etc. Yet, intuitively it is plausible that a person who is creative in one area has neither the time, ability, nor the motivation to be creative in other areas. To date, this hypothesis has been the major focus of only two studies (Hocevar, 1976; Jones, 1964) despite its obvious importance to theoretical and practical concerns. It is almost certain that the personality and intellectual correlates of creativity would be moderated by the area under consideration. In addition, there is the practical problem of predicting creativity
in different areas, since different instruments should probably be used for different areas. Although the need for more research on the role of creativity in different areas has been expressed (McNemar, 1964; Thorndike, 1966), there has been little effort in this direction.

The most condemning problem associated with the measurement of creativity is the lack of convergent validity among different methods. Each of the previously defined approaches is a widely used and accepted technique for identifying creative talent. Since each method is purported to be measuring creativity, it is reasonable to predict that they be correlated, thus satisfying a minimum condition of convergent validity.

In a study of 166 scientists in a government laboratory, Taylor, C., Smith and Ghiselin (1963) collected scores on 52 criteria of creativity including eleven supervisor ratings, six peer ratings, autobiographical information and numerous product variables. While scores based on the same method (e.g., the eleven supervisor ratings) tended to be highly related, scores obtained from different methods were not. For example, supervisor ratings had a low relationship to peer ratings, and self-rating scores were uncorrelated with the more objective autobiographical data. Furthermore, scores based on products such as publications and reports were not related to either supervisor or peer ratings. When factor analyzed, this array of criteria yielded a total of fourteen relatively independent categories.
Davis and Belcher (1971) compared four methods of identifying creativity—two tests of divergent thinking, a biographical inventory, and a criterion which was composed of self-ratings on several creative activity questions. Except for the relationship of the biographical inventory and the criterion, the intercorrelations of these tests were low and generally nonsignificant for both males and females.

In their study of art students, Getzels and Csikzentmihalyi (Note 1) included fourteen measures of creativity, including two teacher ratings, eight tests of divergent thinking, two personality inventories (Cattell's 16 Personality Factors and the Study of Values), grades, and IQ scores. The authors predicted some overlap between these criteria since they all have some claim as indices of creativity. Such was not the case, a factor analysis revealed only method factors rather than a general creativity factor. In another study of art students, Ellison (1973) found low and negative relationships between judged creativity on a pastel drawing and the Remote Associations Test and Barron-Welsh Art Scale.

Finally the correlations of divergent thinking with other measures of creativity have been inconsistent with some investigators finding a positive relationship (Bartlett & Davis, 1974; Bennett, 1973; Cropley, 1972; Dewing, 1970; Halpin, Halpin & Torrance, 1974; Haven, 1965; Hocevar, Note 2; Jones, 1964; Milgram & Milgram, 1976; Rotter, Langland & Berger, 1971; Torrance, 1969a,b;
Wallach & Wing, 1969; Wallbrown & Huelsman, 1975; Zegas, 1976) and other investigators finding no relationship (Andrews, 1975; Barron, 1969, p. 86; Beittel, 1964; Brittain & Beittel, 1964; Dillehunt, 1973; Getzels & Csikszentmihalyi, Note 1; Gcolshy & Helwig, 1975; Gough, 1976; Hadden & Lytton, 1971; Jordon, 1975; Karlins et al., 1969; Kogan & Pankove, 1974; Popperová, 1972; Roweton, Farless, Donham, Wleklinski & Spencer, 1975; Skager, Klein & Schultz, 1967). Further, in studies where significant positive correlations have been reported, the correlation is seldom higher than .30, suggesting that the two measures only have ten percent of the variance in common.

When creativity is identified using multiple methods, individuals are ranked differently. High scorers on one method are not necessarily high scorers on another. This finding raises questions as to whether the trait of creativity is a necessary psychological construct.

Besides these theoretical problems associated with construct validity, the lack of convergent validity has another important implication. Any study of creativity is severely limited by its assessment method. In other words, researchers should limit their conclusions to populations chosen by the same method.

A possible answer to this confusion is that many of the methods which have been tagged with the creativity label do not really measure creativity. At least, they do not measure the behavior that society typically labels creative. In many cases, the methods are
assumed correlates of real life creative behavior. Guilford and the cognitive group have assumed that divergent thinking is somehow linked to creative behavior. The personality psychologists have assumed that some traits are linked to creative behavior. And others have assumed that attitudes and interests or past experiences are linked to creative behavior. Although there is some evidence to support the validity of these assumptions, the relationship of these measures to real life creativity is not one to one. Therefore, their uncritical use as measures of creativity is not warranted.

An important distinction needs to be made concerning the goals of creativity research. It must be emphasized that different researchers have studied different aspects of creativity and that often these cross-purposes have been reflected in the measuring process. Tests of cognition, attitudes, interests, personality, biography, etc. are appropriate when the researcher's goals are to explain something about creativity, but they are not acceptable in the selection process. High scorers on these tests are not necessarily creative people. If a researcher is interested in the personality correlates of creativity or in the relationship of intelligence to creativity, or in evaluating certain educational programs designed to enhance creativity, why not go directly to the observable criteria? This can best be accomplished through studying eminent individuals or using an inventory of creative activities and accomplishments.
Since studies of eminent men are often impractical, the conclusion presented here is that an inventory of creative activities and accomplishments is our best measure of creativity. The kind of activities listed in the Holland and Nichols (1964) inventory are recognized by society as being creative, and they are relatively rare. The activities demand a high degree of commitment and they are not something that just anyone can master.

Furthermore, it is asserted here that a useful way to measure creativity is to simply ask the subject. This is not a profound position, but yet the procedure is rarely used. The predominant preference in the field today is to identify creativity by indirect methods (i.e., predictors) that essentially have little to do with the real criteria of creativity.

Asking the subject has one further advantage. The subject, in most cases, knows more about himself than peers, supervisors, teachers, etc. Earlier it was argued that other individuals cannot always discriminate creativity from their own general opinion of the subject. On the other hand, the subject himself should have a good idea of his creative ability in a wide variety of areas. Furthermore, when compared to observer ratings and other assessment procedures, self-reports have been found to be superior in the measurement of many psychological traits (Mischel, 1968).

An inventory of creative activities and achievements also has predictive utility. Past behavior is generally the best predictor
of future behavior. The efficacy of this predictive strategy is
evident in a number of areas (see Mischel [1968] for a review).
These findings appear to generalize to the area of creative be-
behavior. MacKinnon (1968) reports that, in studies of creative
persons in a variety of fields, earlier accomplishments were con-
sistently predictive of later accomplishments.

The most compelling evidence for predicting future creative
behavior with past creative behavior is found in two related
studies (Holland & Nichols, 1964; Richards, Holland & Lutz, 1967a).
In a high aptitude sample, Holland and Nichols predicted talented,
nonacademic accomplishment in six areas: leadership, science,
dramatic arts, literature, music and art. Predictors included
interests, goals, activities, self-conceptions, aptitudes, and
personality traits. The results of this study indicated that
past activities and achievements were easily the best predictors
of creative achievement in college. The findings of this study
have been replicated with a less select sample by Richards et al.
(1967a).

To conclude, peer nominations, supervisor ratings, teacher
nominations, and judgments of products are often inadequate in-
dicators of creativity due to the rater's inability to discrim-
inate creativity from other traits. Divergent thinking, bio-
ographical characteristics, attitudes and interests, and person-
ality characteristics are best described only as correlates of
real life creative behavior, and they should not be taken as
direct measures of creativity. In addition, most approaches generally fail to discriminate creativity in one area from creativity in another area. Thus, despite the voluminous literature on the measurement of creativity, a simple and straightforward inventory of creative achievement and activities appears to be more defensible than the more commonly used methods.
Reference Notes


Measurement of Creativity

References


Christensen, P. R., Guilford, J. P., Merrifield, P. R., & Wilson, R. C. *Alternate Uses*. Beverley Hills, California: Sheridan Psychological Services, 1960.

Christensen, P. R. Merrifield, P. R., & Guilford, J. P. *Consequences*. Beverley Hills, California: Sheridan Psychological Services, 1958.


Hocevar, D. Studies in the evaluation of tests of divergent thinking: 
(1) Fluency as the confounding factor in the measurement of 
originality; (2) The development of a criterion measure of 
creativity with emphasis on activities and accomplishments; 
and (3) The relationship of ideational fluency and verbal intel-
ligence to creative activities and accomplishments. Unpublished 

Hocevar, D. Dimensionality of Creativity. Psychological Reports, 
1976, 39, 869-870.

Holland, J. L. Some limitations of teacher ratings as predictors 
of creativity. Journal of Educational Psychology, 1959, 50, 
219-223.

Holland, J. L., & Astin, A. W. The prediction of the academic, 
artistic, scientific, and social achievement of undergraduates 
of superior scholastic aptitude. Journal of Educational 

Holland, J. L., & Baird, L. L. The Preconscious Activity Scale: 
The development and validation of an originality measure. 

Holland, J. L., & Nichols, R. Prediction of academic and extra-
curricular achievement in college. Journal of Educational 


Measurement of Creativity

42.


Rivlin, L. G. Creativity and the self-attitudes and sociability of high school students. *Journal of Educational Psychology*, 1959, 50, 147-152.

Roe, A. A psychological study of physical scientists. *Genetic Psychological Monographs, 1951a, 43, 121-235.*

Roe, A. A psychological study of eminent biologists. *Psychological Monographs, 1951b, 65 (no. 331).*


Rookey, T. J. Validation of a creativity test: The 100 students study. *Journal of Creative Behavior, 1974, 8, 211-213.*


Runner Associates. **Runner Studies of Attitudinal Patterns.**


Torrance, E. P. Will creatively gifted high school seniors behave creatively seven years later? TAG Gifted Children Newsletter, 1969b, 12, 24-31.


Wallbrown, F. H. A validity study o. the Wallach-Kogan creativity test: The prediction of six concurrent criteria in visual art. (Doctoral Dissertation, Ohio State University, 1971).

*Dissertation Abstracts International*, 1972, 32, 6630B-6631B. (University Microfilms No. 72-15, 319)


Welsh, G. *Preliminary manual for the Welsh Figure Preference Test*. Palo Alto, California: Consulting Psychologists Press, 1959.

Welsh, G. S. Adjective Check List correlates of intelligence and creativity. Paper presented at the meeting of the Southern Psychological Association, Atlanta, April, 1967. (Yarnell, T. A common item creativity scale for the Adjective Check List. *Psychological Reports*, 1971, 29, 466.)


Yarnell, T. A common item creativity scale for the Adjective Check List. *Psychological Reports*, 1971, 29, 466.