VITA

NAME OF AUTHOR: James Benjamin Lombard

PLACE OF BIRTH: Helena, Montana

DATE OF BIRTH: May 21, 1923

UNDERGRADUATE AND GRADUATE SCHOOLS ATTENDED:

Rocky Mountain College
University of Texas
Columbia Teachers College
University of Oregon

DEGREES AWARDED:

Bachelor of Fine Arts, 1952, University of Texas
Master of Arts, 1953, Columbia Teachers College

AREAS OF SPECIAL INTEREST:

Elementary School Teacher Preparation
Pre-school Education

PROFESSIONAL EXPERIENCE:

Art teacher, Herren Junior High School, La Junta, Colorado, 1953-1954

Art teacher, West Junior High School, Boise, Idaho, 1954-1957

Art teacher, Marshfield Junior High School, Coos Bay, Oregon, 1957-1960

Art teacher, Roseburg High School, Roseburg, Oregon, 1960-1964

Teaching Assistant, School of Architecture and Allied Arts, University of Oregon, Eugene, 1964-1966

Art Instructor, Umpqua Community College, Roseburg, Oregon, 1965-1966

Assistant Professor of Art Education, Southeastern Missouri State College, Cape Girardeau, 1966-1967

Assistant Professor of Art Education, University of Alberta, Edmonton, 1967-1970

PUBLICATIONS:

"Discussion: Teaching Art" (A reaction to a paper delivered by David B. Manzella at the Second Invitational Conference on Elementary Education, Banff, Alberta), The Education of Elementary School Teachers, Robert D. Armstrong and Patricia A. Lane (Eds.), Department of Elementary Education, University of Alberta: Edmonton, 1967, pp. 44-49

"Children's Art Lab at the University of Alberta" (R. N. MacGregor, co-author), Fine: Journal of the Fine Arts Council, Alberta Teachers' Association: Edmonton, Summer, 1969, pp. 57-60

"Art as a Reflection of Man-Reacting-To-His-Environment" Elements: Translating Theory into Practice, Department of Elementary Education, University of Alberta: Edmonton, Vol. I, No. 1, September, 1969

"The Children's Art Laboratory: The Child Sees His World and Expresses What He Sees" (A paper presented at the Conference of Early Childhood Education, Edmonton, Alberta), Conference proceedings to be published in the Spring of 1970 by Early Childhood: Journal of Early Childhood Education, Alberta Teachers' Association: Edmonton

"Individualized Instruction in Art Education? What's New?" (A paper presented at the Third Invitational Conference on Elementary Education, Banff, Alberta), Conference proceedings to be published in the Spring of 1970 by the Department of Elementary Education, University of Alberta: Edmonton

A LONGITUDINAL AND COMPARATIVE STUDY OF 8TH GRADE STUDENTS' "SPONTANEOUS" AND "CREATIVE" ART PERFORMANCE AND PRODUCTION

by

James Benjamin Lombard

A DISSERTATION

Presented to the School of Education
and the Graduate School of the University of Oregon
in partial fulfillment
of the requirements for the degree of
Doctor of Education

June, 1970

APPROVED: _____

Dr. Lloyd Lovell

ACKNOWLEDGMENT

Sincere appreciation is expressed to Dr. Lloyd Lovell and all who offered encouragement and willingly participated during the course of this study.

TABLE OF CONTENTS

LIST OF TABLES

CHAPTER I

INTRODUCTION

> The process of investigation, however, should be unfettered by predetermined research procedures. These assumptions recognize that creative behavior involves complex relationships of variables which now lack clarity. To isolate any variables, therefore, at this time, would be premature and risky. To apply techniques of quantitative analysis, in our present state of knowledge, would open doors to serious prejudgement and error.[1]

This statement appeared in print some ten years ago, and there has been much research into creative behavior in the visual arts in the interim; however, there is no reason to believe, and certainly no irrevocable evidence to indicate, that the statement is much less meaningful today. In fact, it is debatable whether time and investigative efforts have reduced or increased the risk-taking of isolating variables in so nebulous an area as creative behavior in the visual arts. There is a question, too, as to the techniques of quantitative analysis used by some investigators. Most questionable at the present time is the acceptance of inferences drawn from research results as basic criteria for modifying educational practices in art education. In particular instances, these inferences are woven into the

---

[1] M. Barkan and J. Hausman, "Two Pilot Studies for the Purpose of Clarifying Hypotheses for Research into Creative Behavior," Research in Art Education, Seventh Yearbook, Washington, D.C.: National Art Education Association, 1956, p. 127.

authoritarian climate of professional publications and textbooks and have taken on an aura of dicta which should be adhered to in identifying and promoting creativeness in the visual arts. Whether or not this was the intent of the writers is not the question; the underlying implications, real or fanciful, could lead to potentially detrimental practices by unwary readers.

It is a concern that the limits of our present knowledge of creative behavior in the visual arts have been exceeded which motivated this study.

Purposes of the Study

This study has incorporated the use of two evaluative criterion measures which purport to measure, in terms of their authors' defining and normative statements, the degree of creativeness visually manifest in art products. The purposes of this study were 1) to determine whether there are real differences in ranking art products of the same population between judge groups employing these measures, and 2) to analyze any ranking fluctuations of the subjects' art products between and among different art tasks and media as determined by the judge groups' evaluation.

Need for the Study

Researchers in art education have focused mainly on the analysis of art products for manifest visual qualities which characterize

creative products in the visual arts. Some investigators have been intent on developing differential criteria in measuring different aspects of creativeness in the visual arts. Other investigators have pursued a single criterion, claiming their criterion measure to be "global" in nature because it is just as effective as differential criteria for identifying and measuring creativeness in the visual arts. In most research the investigators have either collected art samples for analysis, or they have devised drawing and/or structure tests with predetermined tasks, media, and materials to determine the normativeness of those evaluative measures which are of personal interest to the designers. From a perusal of pertinent literature dealing with this kind of research, there apparently has been little attempt to apply these evaluative measures in an existing, operating educational situation. Nor has there been a comparative study of any of the evaluative measures on the same population. There is a need for an exploratory study which would entail a comparison of the more promising evaluative measures in ranking art products of subjects of the same population over a period of time. Such an extended sampling of art tasks, media, and materials and the evaluation of the resulting art products on the basis of these evaluative scales would indicate the feasibility of the use of these scales for art teachers in evaluating their teaching effectiveness or for identifying potentially gifted students. This study was designed to help answer this need.

Statement of the Problem

This was a five month study of comparing two evaluative criterion scales in the scoring of art product samples of the same population. Both measures are considered by their authors as potentially valuable for identifying and measuring the degree of creativeness reflected by art products and thus, the creativeness of the artists.

One scale, "Six Criterion Elements of Spontaneity," was devised by Gloria Berheim and is a modification and a simplified version of an instrument developed by Kenneth Beittel and R. C. Burkhart.[2,3,4] This is a "global-type" measure whose six elements direct one to different aspects of the art product's surface for identifying the degree of spontaneity of execution in delineating its content (Appendix A). The other scale, "Typology of Creativity," was developed by Elliot Eisner for classifying and defining three levels of creativity within two loci, form and subject.[5] Although this scale is based on

---

[2]Gloria Bernheim, "The Dimensionality of Differential Criteria in the Art Product," Studies in Art Education, 6:31-48, No. 1, (1964).

[3]R. C. Burkhart, Spontaneous and Deliberate Ways of Learning in Art, Scranton: International Textbook Co., 1962.

[4]Kenneth R. Beittel and R. C. Burkhart, "Strategies of Spontaneous, Divergent, and Academic Art Students," Studies in Art Education, 5:20-41, No. 1, (1963).

[5]Elliot Eisner, "A Typology of Creativity in the Visual Arts," Studies in Art Education, 4:11-22, No. 2, (1962).

differential criteria, it is essentially a "global-type" measure because it measures one variable--creativeness--on a continuum (Appendix B). That is, the continuum grows from the simplest novel elaboration of subject matter and forms to the creation of utterly new subject matter and forms.

While the Bernheim scale requires judges to rank art products on the degree of apparent spontaneity in their artistic execution, i.e., lack of contrived delineation of subject and content and the lack of restraint in applying the art media, the Eisner scale directs judges to rank art products on the degree of novel elaboration and combination of their subject matter and forms or the degree of originality in the production of subject matter and form. Essentially, the Bernheim scale is used as a measure of the quality of surface treatment of the art product and the Eisner scale is used to measure the innovative quality of the content and organizational relationships of the art product.

The problem involved the training of three judge groups in the proper ranking procedures required by the two criterion measures. Two three-man judge groups were formed and each was trained in the employment of one of the criterion measures. These two judge groups ranked art product samples. The third judge group was composed of the two teachers involved with the subjects in a "team-teaching" situation. This judge group used both criterion measures, but separately, in judging and ranking the subjects, rather than their art products, at the end of the research period.

The identification of the three judge groups throughout the remainder of this dissertation will be designated by the following shortened titles:

> Six Criterion Judges: The judges who employ the "Six Criterion Elements of Spontaneity" measure for evaluating art product samples.
>
> Typology Judges: The judges who employ the "Typology of Creativity" measure for evaluating art product samples.
>
> Teacher Judges: The two teachers who employ both criterion measures for evaluating their students.

The art products were composed of five sets of two-dimensional art products and three sets of three-dimensional art products and were executed with different art media and materials. The art products were produced over a five month period by 24 eighth grade students who had elected art in a junior high school.

Questions to be Answered

1. Will there be a statistically significant difference between the Typology Judges and the Six Criterion Judges' ranking of the subjects' art products on each art task and across all art tasks?

2. Will the Typology Judges and the Six Criterion Judges arrive at a statistically significant difference of rank-orders among art task pairings?

3. Will there be a statistically significant difference between the pooled rankings of the subjects' two- and three-dimensional art

products? Will this difference be found by the Six Criterion Judges and the Typology Judges?

4. Will there be a statistically significant difference between the pooled rankings of the subjects' art products executed with fluid art media and those executed with inert, resistant art media? Will this difference be found by the Six Criterion Judges and the Typology Judges?

5. Will there be a statistically significant difference between the pooled rankings of the subjects' art products which entailed direct sensorial contact with the art media and those which some intermediary agent (tool) was used? Will this difference be found by the Six Criterion Judges and the Typology Judges?

6. Will there be a statistically significant difference between the Teacher Judges' pooled rank of the subjects and the over-all pooled rank of the subjects' art products by the Six Criterion Judges and the Typology Judges?

Review of Literature

This study entailed the use of art judges trained in the use of criterion scales for evaluation purposes. Many investigators have used judges in their studies, but in the abbreviated reports of their studies in research bulletins very little space is devoted to describing

the time and effort expended in training the judges. In general, the investigators who had an interest in devising and refining scales which used a number of criteria, implied that the training period was long and tedious.[6] There seem to be two dominant reasons why judge training in some studies is so difficult: First, the number and complexity of criteria a single judge is called upon to use makes training periods necessarily lengthy; and secondly, the lack of specificity of the criterion scales seems more the cause of lengthy periods of training than does any ineptness on the part of the judges.

In the case of some researchers who used more simplified criterion scales and who made them operable before the judge training sessions, there are indications that much shorter training periods are sufficient. Eisner reported that a two-week period was required for clarifying his judges' understanding of the criteria which comprised his "Typology of Creativity" scale.[7] On the basis of the actual evaluations, Eisner reports that moderate interjudge reliability was achieved. He reported 42 obtained coefficients which ranged from .90 to .10 with the median coefficient among all indices as being .59 (41 of 42 coefficients were significant at the .01 level of confidence). Frankston

---

[6]For example, Dr. June McFee in an unpublished report on "Creative Problem Solving Abilities of Academically Superior Adolescents" has stated that judges used in this study spent many weeks clarifying the criteria by which they could agree in judgment (p. 13). In an oral communication, Dr. McFee stated that some six months were required to obtain interjudge reliability in this particular study.

[7]Eisner, op. cit., p. 14.

reported the training of judges used in his research to be highly successful in a short period of time.[8] He reported that it took four concentrated sessions to obtain high interjudge reliability (.906) in using a single criterion in judging poetry. On the basis of this experience, he reduced his training sessions to one two-hour session for each of two criteria used for evaluating visual art products. The result was reliability coefficients of .949 and .946.

The lack of specificity of criteria used in evaluation scales can make high interjudge reliability difficult to obtain. "Aesthetic quality" was one of the first criterion used in evaluating creativeness of art products. Burkhart found that judges could rate pictures along a continuum in relation to whether they felt the pictures were aesthetically "good or bad."[9] Each judge could do this reliably; that is, if the judge was asked to repeat the rating of the same art products, he would do it in the same way. However, it was not known what went into these judgments as each judge was left to define for himself what was a good art product. Naturally, such evaluations by a judge would probably not be similar to those of another judge.

---

[8]Leon Frankston, "Some Explorations of the Effect of Creative Visual Art Experiences upon the Poetry Writing Quality of Eighth Grade Students," Studies in Art Education, 5:42-59, No. 1, (1963).

[9]Burkhart, op. cit., passim.

The question that arises is, "What comprises such 'global-type' judgments?" Linderman investigated the relationship between aspects of personality and judgments of art quality by art and non-art personnel.[10] It was shown that judges who are similar in certain personality traits tend to cluster together according to their judge agreement of the same art works. It was also shown that within-group clusters are different from each other, both in personality variables and in art judgment. Beittel did a similar investigation but confined his population to graduate and undergraduate art education majors.[11] His findings were similar to Linderman's in that judgments of a given art product depend upon the kind of personality characteristics of the judge. Furthermore, the kind of aesthetic judgment he makes is highly correlated with the kind of art product he produces, the amount of training he has had, and his level of sophistication.

It can be concluded from the literature dealing with the use of judges in evaluating art products, especially if art teachers are to make the judgments in determining the effectiveness of their instruction or for selecting students for special art programs, that the ideal criteria would be small in number, simple, and easily and logically understood by the judge. This is no simple task, but a far greater

---

[10]Earl Linderman, "The Relation of Art Judgment to Judge Personality," Studies in Art Education, 3:46-51, No. 2, (1962).

[11]Bernheim, op. cit., p. 32.

problem is determining valid criteria for evaluating art products on the basis of their aesthetic and creative worth.

Criteria developed for purposes other than the evaluation of some aspect of the aesthetic or creative quality of art products have enjoyed some success. For example, Karen Machover designed a drawing test as a basis for diagnostic psychological statements.[12] Wolff developed a "Table of Graphic Elements" in order to better evaluate child personality.[13]

The analysis of children's drawings as a basis for measuring intelligence has proved its worth. Goodenough's "Draw-a-Man" test is probably the best known and perhaps the most unusual in basic conception and convenience.[14] Besides being used as a measure of children's I.Q.'s, it has been used as a measure of the extent of children's hearing problems, adjustment problems, and character defects. Several modifications of the Goodenough test have been introduced as "projective tests." Goodenough, in collaboration with Harris, shifted from the analysis of children's drawings as a basis for measuring intelligence

---

[12]Karen Machover, Personality Projection in the Drawings of the Human Figure, Springfield: O. C. Thomas, 1949.

[13]W. Wolff, "Projective Methods for Personality Analysis of Expressive Behavior in Preschool Children," Character and Personality, 10:309-330, No. 4, (1949).

[14]Florence L. Goodenough, Measurement of Intelligence by Drawings, New York: Harcourt, Brace, and World, 1926.

toward greater interest in projective theory, concern over methodological questions and more exact measurements, and the usage of children's drawings in solving problems in general psychology.[15] Recently, Harris has revised and attempted an extension of Florence Goodenough's original scale.[16]

In the field of art education, attempts to identify critical aesthetic and creative criteria for the purpose of evaluating art products have greatly multiplied in the past few years. To date, art educationists are not in agreement as to which criteria are essential for the evaluating of aesthetic and creative qualities of art products. Some of the scales which have been developed hold promise, but, in spite of their authors' tentative avowals of possible worthiness as evaluative devices, they have yet to be thoroughly tested.

In the past, and even up to the present time, juries of "art experts" have been used to evaluate art works. At its worst, the employment of this procedure has left it up to the individual members of the jury to determine what criteria to use in evaluating art works. At its "best," the members of a jury have reached, before-hand, a consensus of opinion as to what criteria are essential for determining

---

[15]Florence L. Goodenough and Dale B. Harris, "Studies in the Psychology of Children's Drawings: II, 1928-1949," Psychology Bulletin, 47:369-433, 1950.

[16]Dale B. Harris, Children's Drawings as Measures of Intellectual Maturity, New York: Harcourt, Brace, and World, 1963.

the value of individual art products. The latter example is an inductive method; that is, the jury evaluates only those characteristics which they have decided are most essential to "works of art." Whether evaluative criteria are determined by the individual judge or by group consensus, there are a number of problems inherent in the art jury system.

Assuming the unlikely possibility that individual members of an art jury possess the essential criteria for evaluating art products, it is likely that there will not be a high degree of consensus among the judges as to the rank value of individual art works. One way which is often used to resolve these differences is to take the mean average of the judging results. This kind of averaging may not reflect the "true" rank value of an art work. As an illustration, an art jury of three judges using the rank-order method of scoring a set of art works arrive at the following "scores" for two of the art works:

	Judge 1	Judge 2	Judge 3
Art work 'A'	1	1	4
Art work 'B'	2	2	1

If the mean average of the ranks was used to determine the rank value of individual art works, art work 'B' would be considered better than art work 'A.' However, the majority of the judges (judges 1 and 2) were in agreement that art work 'A' was better than art work 'B.' In this case, the median score rather than the mean average could be more

indicative of the "true" rank value of the art work. It is well to bear in mind, however, that interjudge agreement is no assurance that the criteria used in the evaluation of art works are valid. Nor are the criteria used by a judge who is at odds with the rest of the jury necessarily less valid than the criteria employed by the majority of the members of an art jury. Furthermore, the difference of ranks between judges can be far greater than shown in the illustrative example and, thus, result in still greater misjudgments as to rank values of individual art works.

Another way of resolving score differences is for the judges to arrive at a consensus as to the rank value of those art works in question. Essentially, this is the same procedure used by members of an art jury who have reached, before-hand, a consensus as to what criteria should be used for determining the value of individual art works. One of the problems with this procedure is that the resulting "compromise" is influenced by the forcefulness of the individual personalities of the judges and their individual commitment to the judging task. Another problem is that by determining ahead of time what criteria should be attended to for evaluating art works, other attributes of the art works are eliminated from consideration during the judging procedure. These other attributes may be essential to the purposes of particular art judging tasks.

It is reasonable to assume that individual members of an art jury have generally employed inductive reasoning to arrive at their

personal criteria for evaluating art works; that is, individual judges draw upon their past experiences for values which they feel are essential for measuring the worth of art works. From this valuing, inferences and conclusions are drawn which serve as the bases for defining their personal art valuing criteria. The inducted criteria will determine, to a very real extent, what aspects of art works a judge will attend to in his evaluations, and as pointed out previously, if a judge's criteria for art judgment are limited in scope or not valid for evaluating a particular set of art works, serious misjudgments can result.

Within the defining structure of research, most searches for the critical aesthetic and creative criteria for evaluating art products have employed the inductive method of investigations. Researchers in art education have relied upon unstructured observations from their experiental background or have relied upon "facts" derived from research results in other disciplines as a basis for determining criteria to be investigated. Researchers have also used "art experts," in the sense that art judges have been trained in the employment of the criteria which they have refined and clarified for their studies. Furthermore, the magnitude and significance of interjudge reliability plays no small part in this method of investigation.

It is well to reiterate a possible danger in accepting results from this kind of research too readily. A high or significant correlation between judges does not mean that their judgments are correct. It is possible that a number of judges can agree in their judgments

because all employ the "wrong" criterion. As Siegal points out, "objective" ratings and "consensual" ratings are not necessarily synonymous.[17] However, until research methods become more refined, researchers in art education must continue to rely to some extent on this kind of research in their searches for critical criteria for evaluating art products.

Too little art education research has incorporated the deductive method in conducting investigations into criteria for evaluating art products. There is a need for more research which would be in the nature of objective description; that is, an inventory of the visual characteristics and their interrelatedness as manifested in art works. Furthermore, such a definitive inventory must be precise enough to assure the highest degree of agreement possible among informed observers. This kind of research must be done before any interpretation or valuing of art-product characteristics can be undertaken for identifying essential aesthetic or creative qualities of works of art. It is understood that these qualities are not analyzable properties in themselves; they exist only in relation to something other than themselves, so it becomes important that all related variables are also precisely described. Essentially, the foregoing comments discriminate between descriptive criteria and those criteria which may be identified

---

[17]Sidney Siegal, Nonparametric Statistics for the Behavioral Sciences, New York: McGraw Hill Book Company, Inc., 1956, p. 238.

from among descriptive criteria as essential for a particular kind of evaluation of art works.

Too few studies have attempted to determine what kind of instructional experiences would contribute to the development of creative traits. Cramer Owen achieved some success in his two-and-one-half year longitudinal study designed to stimulate original thinking.[18] June McFee, using standardized creativity tests and personality measures that discriminate between more and less creative traits, explored the relationship of a course of study, including problem-solving in design and the analysis of creative behavior, to the creative development of academically superior adolescents.[19] The experiment was conducted during 1960-61, and the conclusion reached was that the particular type of training used during this experiment was effective in changing the responses of students of this ability level and general socio-economic level to the particular measures of the behaviors subsumed under the concept "creativity."

In the search for essential criteria for evaluating the aesthetic and creative qualities of art products, there has developed a division of opinion as to whether a single criterion is just as effective as multiple criteria in evaluating art products. Lansing concluded that

---

[18]Cramer Owen, "An Investigation of Creative Potential at the Junior High Level," Studies in Art Education, 3:16-33, No. 2, (1962).

[19]McFee, op. cit., passim.

a single judgment of children's drawings is as good as several criteria judgments.[20] Paul Edmonston insists that evaluation of art products must depend on the various purposes of the evaluation and upon multiple criteria.[21] Recently, Mary Rouse conducted an interesting study which attempted to avoid this global-type evaluative issue of "goodness" or "badness" of art products by seeking to identify and utilize, by deductive means rather than inductive means, characteristics which might be termed "descriptive."[22] The research procedure used by Rouse is considered deductive because she began with the basic assumption that the defining terminology which serves as a basis for evaluating art works is, in effect, identifiable in common usage. Surveying art literature, Rouse identified and deducted commonly used descriptive terms. The accumulated list of terms was delimited and the final terms operationally defined. The final list was submitted to 48 experts

---

[20]Kenneth Lansing, "The Effect of Class Size and Room Size upon the Creative Drawings of Fifth Grade Children," Research in Art Education, Ninth Yearbook, National Art Educational Association, 1959, pp. 70-79.

[21]Paul Edmonston, "Overview: Product and Evaluation," Art Education Bulletin, Kutztown: Eastern Arts Association, 18:51-59, No. 4, 1961.

[22]Mary Rouse, "The Development and Validation of a Descriptive Scale for Measurement of Art Products," Cooperative Research Project No. S-077, Cooperative Research Program of the Office of Education, Department of Health Education, and Welfare, (An unpublished study conducted at Indiana University, Bloomington, Indiana), 1956.

in the field of art who used Q-sort to rank the efficiency of the descriptive terms for the purposes specified. Rouse felt that the final scale made possible more precise descriptions of behavioral changes.

There have been several reports of studies pertaining to the two "global-type" criterion scales which are pertinent to this study. Burkhart isolated and synthesized attributes which dichotomized art students into spontaneous-deliberate modes of art work.[23] By his definition, this dichotomy was equated with the presence or lack of creativity. The interest in the spontaneity criterion was probably stimulated by the outpouring of research results on creativity which was so prevalent at that particular time. Such research results pointed to the attributes of fluency and flexibility as being most characteristic of highly creative people. Also spontaneity is a dominant characteristic of Abstract Expressionism which was the dominant style of painting in the United States at the time of the Burkhart study. If these influences were not directly responsible for the induction of the criterion of spontaneity, they probably entered into the thinking of the researchers who pursued the development of this criterion.

Research into the criterion of spontaneity has been advanced by Beittel and Burkhart.[24] Assigning 47 college junior art education

---

[23]Burkhart, op. cit., passim.

[24]Beittel and Burkhart, op. cit., passim.

majors painting problems on defined tasks, they closely observed the work strategies employed by their subjects by using photographs taken at three minute intervals as the work developed. Tests had been given to the subjects four weeks prior to the studio learning experiment, and these tests were considred by the authors as predictive of the work strategies employed by their subjects in their later art works. The tests were devised by the authors and others for the specific purposes of this study.

The authors' procedure for identifying work strategies and their choice of tests has been criticized by Chapman.

> Contrary to one's first impression then, the strategies do not appear to have been initially identified by drawings at all. Indeed the very makeup of the experimental groups in the studio learning experiment does show that the strategy distinctions were clearly established before the drawings were even made. "Strategies" then, are interpreted from an analysis of test scores, rather than from the art products per se. Since the test scores are the primary basis for distinguishing among strategy types, it would be reasonable to ask whether the tests are standardized and, if they are not, on what grounds the authors reject standardized personality and other tests as useful for their research purposes.[25]

Beittel in reply to Chapman's criticism admitted that more work needed to be done on the tests, and as to whether the authors interpreted "strategies" from art product judgments or the tests, there was no reply.[26]

---

[25]Laura H. Chapman, "Some Comments on 'Spontaneous, Divergent, and Academic Art Students,'" Studies in Art Education, 6:25-29, No. 1, (1964), p. 27.

[26]Kenneth R. Beittel, "In Reply to Laura Chapman's Critique," Studies in Art Education, 6:30, No. 1, (1964), p. 30.

Description of the tests used in this study may be found in Appendix C.

Of the 47 subjects, the authors classified 15 as being Spontaneous, 15 as Divergent, and 17 as Academic. The authors proceeded to analyze how different strategies manifest themselves in art products and, also, how these strategies are based on clear differences in the students' personality structures and in their problem-solving, discovery, and evaluative powers, suggesting that there are operational systems cutting across disciplines. Thus, they constitute different life orientations whose etiology would be worth understanding for educational purposes.

Although significant correlations were obtained between spontaneity and aesthetic judgments and creative personality measures, there is a question whether the spontaneity criterion is an adequate measure of artistic ability. The authors noted the possibility that the student whom they classified as "divergent" was possibly being overlooked because his work lacked spontaneity.[27] Furthermore, Beittel, in his reply to Chapman's criticism, said, ". . . we have abandoned the academic strategy as untenable or as equivalent to the null class of dynamic strategies."[28] From this comment, it can be assumed that the

---

[27]The "divergent" student does creative art work but is not considered spontaneous in his work strategy. In this report the Burkhart deliberate-spontaneous continuum has been trichotomized by breaking down the deliberate classification into a divergent-academic division.

[28]Beittel, op. cit., p. 30.

authors feel that the divergent and spontaneous strategies are still tenable.

Barkan's theory would seem to support Beittel and Burkhart's classification of the "divergent" student when he suggests that it might be possible to place individuals working in the visual arts on a "concept-bound--percept-bound continuum."[29] The concept-bound (subject oriented) has a clear idea of what he is going to do. Such a person tests his art experience against previously held expectations--images he has fixed. The percept-bound individual (form oriented) is more flexible and often takes his cues from emerging visual stimuli. Lantz has also taken note of this difference in work strategies as employed by students at the elementary school level.[30]

Beittel and Burkhart draw some conclusions and formed some hypotheses from their investigations which stimulated the initiation of the present study. For one thing, they pointed out that the possession of a strategy is no assurance of quality. That is, the kind of strategy used in making an art work would be manifest and identifiable in the art product, and that the quality of the art production would not operate as an intervening variable in the identification and measurement of the employed strategy. If this is so, then strategies should

---

[29]Barkan and Hausman, op. cit., pp. 126-141.

[30]Beatrice Lantz, Easel Age Scale, Los Angeles: California Test Bureau, 1966.

be identifiable in the art productions of children who are characteristically less knowledgeable, sophisticated, and skilled in art making than the college-aged subjects who were used by Beittel and Burkhart as subjects in their investigations. The question is whether children have developed and consistently employ a particular strategy in making their art productions.

Secondly, Beittel and Burkhart emphasize throughout their reporting the innovative quality of art works of the subjects who employ the spontaneous work strategy and that this working procedure is characterized by discovery and invention. Of interest, is whether there is a significant linear relationship between the degree of exploitation of a work strategy as manifest in children's art works and other art scales which purport to measure the degree of creativeness of children's art products.

The authors pointed out the possibility that different content areas may require one strategy more than another; that is, in disciplines outside the area of the visual arts, it may be found that one strategy for learning within a particular discipline is more applicable than another learning strategy because of the nature of the subject matter of that discipline. However, the possibility that different content areas <u>within</u> the visual arts may require one strategy for learning more than another was not suggested by Beittel and Burkhart. It seems reasonable that the art materials, processes, and skills identified with a particular art form may require one strategy whereas

another strategy would be more applicable to a different art form. Furthermore, if a child has developed and relies on a particular strategy for art making in all art forms, then his level and consistency of creative work may only be manifested in those content areas in which his particular learning strategy is most applicable. In the present study, the subjects' art products were dichotomized in three ways to test this possibility (two- and three-dimensional art products, fluid and inert art media, and direct and indirect contact with the media).

Gloria Bernheim has attempted to give Burkhart's global criterion of spontaneity greater specificity by investigating what elements comprise this criterion.[31] Although the six elements of spontaneity developed by Bernheim were different from the Burkhart scale, they were found to correlate almost perfectly (.95) with the Burkhart scale, thus giving her scale almost equal reliability as the Burkhart scale. Due to the greater specificity of Bernheim's differential elements of spontaneity, it was decided to use this scale in this study.

The second "global-type" criterion scale used in this study was developed by Eisner.[32] His "Typology of Creative Behavior" consists of classifying and defining three levels of creativeness within two loci, form and subject. That there are different degrees, kinds, or levels of creativeness is not particularly novel; however, the use of the

---

[31]Bernheim, op. cit., passim.

[32]Eisner, op. cit., passim.

concepts of degrees, kinds, or levels of creativeness in the evaluation of art products is novel. Anne Roe grades artists in different occupations and classifies them into five levels of creativity according to their occupations.[33] Maslow defines four kinds of creativity: primary, secondary, integrated, and peak experiences, and sees personality rather than achievement (product) as the crux of the research problem on creativity.[34] Dubin's typology of deviant behavior lists and defines fourteen types of deviant behavior of which five would be considered as definitions of creativities.[35] This typology is an extension of a model of deviant behavior from a sociological point of view. The original model was devised by Robert Merton and contained four types of deviant behavior.[36]

Eisner, in developing his "Typology of Creativity," used eighty-five pre-adolescent subjects who, in isolation, were given a drawing

---

[33]Ann Roe, The Psychology of Occupations, New York: John Wiley and Sons, 1956, pp. 236-38.

[34]A. H. Maslow, "A General Theory of Personality Research," An abstract from a speech (via radio) at the Interdisciplinary Symposium, Michigan State University, 1959.

[35]Robert Dubin, "Deviant Behavior and Social Structure: Continuities in Social Theory," American Sociological Review, 24:147-164, No. 2, (1959).

[36]Robert K. Merton, Social Theory and Social Structure, Glencoe, Illinois: Free Press, 1957.

and a building test which were evaluated by three judges trained in the use of the evaluative criterion. Two of the major purposes of the study were to test the usefulness of the typology and to determine the relationships existing among the various types of creativity.

To determine the utility of the instrument, Eisner relied on the extent of interjudge agreement. It was reported previously in discussing judge training, that Eisner considered interjudge reliability sufficiently high to warrant the usefulness of the measure.

To determine the relationships among types of creativity, raw scores were transformed to 'T' scores and summed, and all indices were then intercorrelated and factor analyzed. High correlations were found among types within drawings but not within the structure task. Eisner points out that the nature of the drawing task may have made it exceedingly difficult for the judges to make discrete judgments on the drawing products. The nine-page booklet used in the drawing task contained abstract lines which served as stimuli for the drawings, and this may have limited the maximizing of differences between types. In the present study, the ranking procedure used in evaluating art products and the greater difference of delineations among art products should overcome this problem.

Of interest to the present study are the low correlations Eisner obtained between creativity expressed in different media. These lower correlations indicate that correlation coefficients drop when different sensory modalities are solicited. The possibility derived from the

Beittel and Burkhart study that different content areas may require one learning strategy more than another may have relevance to the Eisner finding.

Both the spontaneity measure and the Typology of Creativity were developed primarily by inductive means; that is, the criterion measures were seemingly initiated on the basis of theory and research findings in the behavioral sciences and/or the predominant art philosophy of the time. Inductive reasoning as a means of initiating a criterion which may potentially be of value for evaluating the artistic or creative worth of art products is sound. However, and as it was pointed out previously, a well-trained art judge will attend to only those aspects of an art work which are within the operational definition of the criterion, and the possibility exists that serious misjudgments could result if the criterion measure is limited in its scope of measurement.

Both measures are more than descriptive as the criterion in each case was singled out and valued as essential for measuring the creativeness of art products. It is true that both criterion measures could be used for describing art products but the description would be limited to those characteristics which were established before-hand and which fall within the operational limitations of the definitions of the two criterion measures.

The use of creative or spontaneous judgments of art products may have other serious limitations. In the case of evaluating art

products for the early identification of the potentially gifted in art, it may well be that such judgments may tell more about students' developmental levels in art than indicating artistic potential. The use of these two criteria for evaluating the effectiveness of a certain method of art instruction could possibly lead to misleading and erroneous conclusions due to the past kinds and amount of art experiences of the subjects. This consideration is especially important and unique in art education research. Art education, unlike instruction in most school subjects, is not a continuous process in most children's education. Junior high school students' art experiences, for instance, could possibly range from none to a continuous progression of art instruction dating from pre-school days.

Another possible limitation of the criterion Spontaneity is indicated by a study by Jean Holland.[37] She points out that the "accidental" and the "naive" seem to influence art judgments of both art and non-art people. In this regard, such described elements of Spontaneity as "blurred and rough contour," "loose and free forms," and "flexibility in the treatment of the whole" could possibly be part of naive art works. Furthermore, familiarity with a particular medium and/or process could very well be a crucial variable in determining an art product's position on the deliberate-spontaneous continuum.

---

[37]Jean Holland, "A Developmental Scale for Grades 6, 7, 8," (Unpublished Research, Toronto School System, Toronto, Canada, 1958).

There are several research studies and theories which may bear on the problems encountered by the foregoing researchers. The variance that Eisner noted between creative types within different media and Beitell and Burkhart's non-spontaneous "divergent" art students could be accounted for by how they orient themselves to space. Lowenfield's concept of a "visual" and "haptic" continuum, Witkin's classification of "posturally oriented" and "visually oriented" types, and the significant findings of studies concerning learning and personality dynamics in perception could serve as a theoretical basis for art student preference for certain art media.[38, 39] For instance, a student who depends on body cues (posturally oriented) may be stimulated by the fluid quality of watercolor. His art products may be considered more creative in this medium than in other media. It should also be recognized that the preference of a medium and the creative quality of art products in this particular medium may be due to his understanding the potentialities of the medium.

Implications from Readings

In the review of research reports which have incorporated the use of judges and their training in the use of certain evaluative

---

[38]Viktor Lowenfeld, Creative and Mental Growth, 3rd ed., New York: The Macmillan Company, 1957.

[39]Herman Witkinn, "Perception of the Upright," Scientific American, (February, 1959), pp. 50-56.

criteria, the difficulties encountered and the length of time required in training the judges seem to be contingent upon (1) the number of criteria a judge is called upon to use, (2) the complexity of the criteria, and (3) the degree of specificity of the criteria definitions. In studies in which judges were called upon to make their evaluations on the basis of a small number of criteria and when these criteria were well-clarified before the beginning of the training session, relatively short training periods were needed. In the case of one study only one concentrated two-hour session was needed for each of two criteria used in order to obtain high interjudge correlations.

The implication is that in the use of judges in evaluating art products, and especially if teachers are to serve as judges in determining the effectiveness of their instruction or for selecting students for special art programs, the ideal criteria used in their evaluations would be small in number, simply defined, and easily and logically understandable.

The problem in the evaluation of art products can be stated in the form of two questions: What are the essential criteria for judging art products? Is a single criterion as effective as multiple criteria in making art judgments?

The evaluation of art products for diagnostic purposes and for measuring intelligence has enjoyed some success; but in the determining of aesthetic and creative qualities of art products, there is little consensus of opinion. Many criteria scales dealing with the aesthetic

and creative qualities of art products have been devised by both inductive and deductive means. Some are purely descriptive; others claim to measure the essential attributes of art work.

Too few studies have attempted to determine the effectiveness of certain kinds of art instruction.

Many of the studies have used the method of collecting children's art products for analysis and devising evaluative scales. But there is no indication in art education literature that these scales have been applied for the purpose of evaluating different art task products, nor have they been used in longitudinal studies.

Some evaluative scales, although different in terminology and definitions, purport to measure the same qualities of art products. This is the case with the two evaluative scales pertinent to this study. Eisner's "Typology of Creative Behavior" and the criterion "Spontaneity," which is the result of research by Burkhart, Beittel, and Bernheim, purport to measure, by the authors' own definition, the creativeness of the art product. There has been no comparative study to date as to whether art product ratings as determined by judges using the two measures would be similar.

An issue which is not yet resolved is whether a single criterion is as effective as multiple criteria for evaluating the aesthetic and creative qualities of art products. The resolving of this issue would have important implications on the ability of art teachers to effectively

evaluate their art instruction. The need to use multiple criteria for such evaluation could make such evaluations exceedingly difficult because of time and effort.

Researchers who have devised single criterion measures suggest the possible effectiveness of their measures in rating art products in spite of the increasing evidence that other variables could influence the ratings of subjects' art products in any specific instance and may result in a great deal of fluctuation of the rating of subjects' art products over a period of time. The amount and kind of art experiences, familiarity with the art medium, assigned task, personality factors, environmental setting, socio-economic background, and the learning and personality dynamics in perception are but a few variables which could influence the art product rating at any one time or from time to time.

It is well worth reiterating that the possibility of using a single criterion for rating art products may be premature at this time, if not entirely infeasible.

CHAPTER II

PROCEDURAL DESIGN OF STUDY

This chapter describes the procedural steps in conducting this study and the underlying rationale of its organizational programming. More specifically, the following four aspects will be considered: 1) hypotheses, definitions, and limitation, 2) population and sample, 3) collection and presentation of the art products, and 4) the statistics used in the study.

Hypotheses

The following statistically significant differences are predicted:

1. There will be a difference between the pooled rankings of the Typology Judges and the Six Criterion Judges on each art task and across art tasks.

2. There will be a difference between the pooled rankings of the Typology Judges and the Six Criterion Judges among art tasks.

If Hypothesis 2 is accepted, the following Secondary Hypotheses will be tested:

3. There will be a difference between the pooled rankings of the subjects' two- and three-dimensional art products. This difference will be found between the pooled rankings of each judge group.

4. There will be a difference between the pooled rankings of the subjects' art products executed with fluid art media and those executed with inert art media. This difference will be found between the pooled rankings of each judge group.

5. There will be a difference between the pooled rankings of the subjects' art products which entailed indirect sensorial contact with the art media and those which entail direct sensorial contact with the art media. This difference will be found between the pooled rankings of each judge group.

If any one or more of Hypotheses 3, 4, and 5 are confirmed, further differences will be looked for between and within the two judge groups' ranking of products of the art tasks.

In comparing the results of the Teacher Judges to those of the Typology Judges and/or the Six Criterion Judges, the following difference is predicted:

6. There will be a difference between the Teacher Judges' pooled ranking of the subjects and the ranking of the subjects by the Typology Judges and the Six Criterion Judges. The ranking of subjects by the Typology Judges and the Six Criterion Judges will be determined by the pooled rankings of the subjects' art products.

Definitions

Spontaneity and creativity are specifically defined by the two criterion measures used in this study. In order to differentiate these definitions from other possible meanings, a special designation is needed to identify these terms. Such a designation is needed to eliminate confusion and reduce redundancy of definitions.

"Spontaneity." This term is defined by the "Six Criterion Elements of Spontaneity" which comprise the spontaneity domain developed by Bernheim (Appendix A). Basically, it is an evaluative scale used for determining the degree of freedom for innovation during the execution of an art work as judged from the finished art work. Quotation marks enclosing the term "spontaneity" will be used throughout this dissertation to identify its relationship to this evaluation scale.

"Creativity." This term is defined by the "Typology of Creative Behavior" developed by Eisner (Appendix B). The "Typology of Creative Behavior" is a scale devised for identifying and measuring various types of creativity found in visual arts products. Quotation marks enclosing the terms "creativity" or "creativeness" will be used throughout this dissertation to identify their relationship to this evaluation scale.

The following terms and their definitions are used in the testing of the hypotheses of this study:

Two-Dimensional. This term refers to those art activities pursued in art classes which begin with the application of media to a

two-dimensional surface and retains the essence of the two-dimensional quality. The third dimension of depth is found only in that fraction of an inch which measures, for example, the thickness of the paper or paint layer.

Three-Dimensional. This term refers to art products which possess any space-enclosing characteristics or mass forms which deviate from the essentially two-dimensional qualities. This would include products whose materials are inherently three-dimensional or those materials that are two-dimensional but are used to make sculptural-like forms. Examples are salt rising diaramas and paper sculpture.

Fluid Art Medium. This term refers to any art medium which in its working state lacks any significant degree of resistance to a subject's manipulation. Watercolors and clay are examples.

Inert-Art Medium. This term refers to any art medium that requires effort to modify and does not lend itself easily to corrections. Plaster of Paris and linoleum block printing are examples.

Direct Sensorial Contact. This term is used to describe how an art medium is manipulated in the execution of an art product. It is defined as the bodily contact with the art medium in which there is direct bodily sensation with the medium. Fingerpainting and clay modeling with only the hands are examples.

Indirect Sensorial Contact. This term refers to the use of an intermediate agent (tool) by a subject when manipulating and modifying an art medium in the execution of an art product. Examples are painting with a brush or carving with a knife.

Limitations of the Study

1. Subject and art product measurements will be limited to the two criterion measures identified in this study. Other criteria and criterion elements which the authors of the two criterion measures attempted to isolate and develop, but which proved questionable, were not incorporated in this study.

2. The population from which the sample of subjects was drawn had an appearance of heterogeneity in terms of socio-economic and scholarship ability. There were no detailed or comprehensive data on the population which was available to this investigator. Heterogeneity was determined from generalized statements of the school's staff and observations of the neighborhoods from which this school drew its population.

3. The sample of art students drawn from this population consisted of 24 eighth grade students. The subjects were at the end of their schooling in which formalized art training had been minimal and in the early stage of their schooling in which art instruction from qualified teachers was being experienced.

4. The subjects were taught by the same art teachers.

5. The subjects were enrolled in a continuous, two-semester art program.

6. The subjects were not homogeneously grouped by either artistic or mental abilities.

These limitations were the result of the underlying rationale of this study. Basically, the research design was structured so the two evaluative criterion scales could be applied and compared in an operating art learning situation. These evaluative criterion scales, "Six Criterion Elements of Spontaneity" and the "Typology of Creativity," are assumed to have an underlying, continuous distribution in the art products of the population under study. This was the only assumption made in structuring this study and in choosing the statistical tests which were applied to the analysis of data. This assumption and its relationship to choice of statistical tests will be more fully discussed later in this chapter under the sub-heading, Statistical Analysis.

One of the primary purposes of the present study was to determine if there would be ranking fluctuations of the subjects' art product ranks between and among art tasks and media, and, if such fluctuations did exist, to analyze these fluctuations for systematic variables which may be related to particular kinds of art tasks and media. The incorporation of any criterion elements not completely developed by the authors of the two criterion measures would have been of no particular value in the pursuance of this purpose. Furthermore, in pursuance of this purpose, certain modifications had to be made of the two criterion scales.

In the case of Eisner's "Typology of Creativity," the scale was limited to that portion which would lend itself to the study of one variable on a continuum. His types of creative behavior identified as boundary breaking, inventing, and boundary pushing were seen as amenable to a continuum. It was pointed out in Chapter I that Eisner was not able to identify discrete types on his drawing task, and it was posited that ranking art products on a continuum should overcome this problem. Eisner's creative behavioral type "aesthetic organizing" was deleted since it had no position within the continuum of the other creative behavior types; aesthetic organizing was seen as more truly a type or kind of creative behavior in the visual arts, rather than a level of creative behavior attainable on a continuum. Aesthetic organizing forms its own continuum and could be used for judging art products and for comparison with other art scales which deal with aesthetic organizing. Bernheim, for instance, in developing her "Six Criterion Elements of Spontaneity" was trying to identify elements which would comprise other domains similar to the "Spontaneity Domain." Her "Formal Domain" (or "Aesthetic Domain") was less than successful and was not included in the present study. This gave added reason for not including an aesthetic variable for comparison between the two criterion measures. Another "domain" Bernheim was attempting to develop as a basis for global judgments of art products was the "Divergent Domain." It, too, proved not to be successfully determined and is not included in the present study.

As with Eisner's "Typology of Creativity," Bernheim's "Six Criterion Elements of Spontaneity" was modified from a scoring scale to a ranking scale. "Spontaneity" was seen to exist on a continuum and amenable to rank-ordering of art products. Furthermore, there was the question whether the sum of scores derived from the scores obtained on the six elements would give the "true value" of an art product in comparison to other art products; that is, it seemed possible that a particular art product would not have all elements of spontaneity present and yet would be much more spontaneously executed than an art product with all elements present.

Within the framework of the stated limitations of this study there was an attempt to attain a typical population and sample. The primary purpose was to obtain a spread of abilities as measured by the criterion used by the judges in order to 1) assure maximum opportunity for determining the relatedness of the two criterion measures, and 2) ascertain if any of the subjects displayed a consistent level of creativity in their art products as defined and measured by either or both of the evaluative criterion used by the judges in this study. However, this was to be done without interfering with the normal classroom procedure. Such interference could have led to undesirably biased and distorted data. The need was to find an "average" art classroom operating under "normal" conditions.

There was no delusion in thinking that average or normal, as used in this sense, could be adequately defined, nor would it be

possible to find such an ideal situation if these terms could be defined.

With these limitations in mind, the average classroom is defined as one in which the subjects are representative of the larger school population which, in turn, appears to be heterogeneous in terms of socio-economic and scholarship characteristics. Such an art classroom would also be "average" in terms of the scope and sequence of art instruction in art classes at this educational level. Averageness is more speculative than factual in this regard. Although it is relatively common to find a Master's thesis which, in a descriptive nature, analyzes secondary school programs in art at the local level, there have been few studies in this area at the national level. There was a small outpouring of this kind of research after the 1955 National Art Education Association Convention when concern was voiced over junior high school art programs and the need to develop a clearer idea of the then current directions in art curriculum and programming. Two studies of this period were national in scope.[1,2] Both studies

---

[1]Amy G. Brainard, An Analysis of Art Programs in Selected Junior High Schools in the United States, Doctoral Dissertation, University of Southern California, 1955.

[2]Helen C. Rose, "Directions in Junior High School Art Education: A Pilot Survey of City Junior High School Art Programs," Ninth Yearbook, Kutztown, Pennsylvania: National Art Education Association, 1959.

indicated that there were great differences in the organization and presentation of art in the junior high schools; however, breadth of art materials, processes and activities were quite evident. The Helen Rose study included art educators' suggestions as to what emphasis there should be for improving the art program in junior highs.[3] In sum, the art educators advocated still greater breadth of art activities, including close relationship to other school and community needs, experiences and subjects; wider exploration of the many fields of arts and crafts; and relating art to dress, grooming and everyday learning. Although this was some ten years ago, and in spite of a growing emphasis today to teach in greater depth in some areas of art, this emphasis on breadth of art materials, processes and activities is probably the most common characteristic of junior high school programs today.

Normal conditions refer to the usual art offerings and the usual means of presentation of these offerings which the art instructors of this class would normally exhibit.

The study called for a group of subjects who had little formal art training but who were at the time of the study under the tutelage of a qualified art teacher. The need was for subjects who would display art behavior which had not been sophisticated or otherwise modified by art training to any great extent. In this sense, they

---

[3] Ibid., p. 135.

would be displaying their "natural" developmental level of ability for self-expression and creativeness in the visual arts. Of course, cultural and environmental background could not be controlled. It was also realized that during the course of the study their "natural" art behavior would be modified, but if the subjects were experiencing their first formal art activities, it is doubtful that the five-month period would do much in modifying their behavior to any great extent.

A qualified art teacher was a requisite because he or she would be called upon to use the two evaluative criteria. Although it may be posited that the only requisite judge qualification is the ability to use the art judging criteria with a similar degree of confidence as other judges using the same criteria, it was thought best to use trained art personnel to obviate any unnecessary criticism. A qualified art teacher would be considered as a person trained in art education and certified by the state to teach art.

In order to insure that all subjects would experience art activities in the same psychological environment during the time period of the study, they would be drawn from the same school and class and taught by the same teacher. Furthermore, to maintain "average" art class characteristics of this educational level, no group of subjects would be selected who had been grouped according to any artistic or mental abilities.

In order to conduct a five-month study a school which offered a continuing two-semester course of art was necessary.

Finally, the subjects selected for the study had to have, in the course of the study, experience in both two- and three-dimensional art work to insure adequate opportunity to evaluate art products from a breadth of art activities.

Population, Sample, and Art Program

Eugene, Oregon

Eugene is located at the southern end of the Willamette Valley, and including the neighboring city of Springfield and smaller communities, the population of the immediate area is well over 100,000. The main industry of the area is lumber, but in general, industry is quite diversified with agriculture and tourism prominent. Eugene is also the location of the University of Oregon, and in cultural opportunities it compares quite favorably with other cities of similar size in the United States.

Wilson Junior High School

Eugene Public Schools are organized on the 6-3-3 plan. Wilson Junior High School houses some 700 students and is located in the southwest part of the city. The students who attend this school come from homes of varied socio-economic backgrounds. To the southeast the school boundary is adjacent to the university area and many of the students from this area are children of university staff and students. To the northeast and northwest the school boundaries include the

downtown area and the immediate surrounding area. Many of the students from this area are children of blue-collar and white-collar workers. Within the south and southwest boundaries the students are from professional, business-owning, and white-collar families.

Art at the junior high

As with many junior high schools the opportunity to elect courses is limited; one elected course is standard in the seventh and eighth grade and two electives in the ninth grade. Art, band, and choir are offered in the seventh grade. In addition to these three electives, typing is offered in the eighth grade. In the ninth grade a foreign language can be elected by some students. Art is offered on a semester or two-semester basis and is arts and crafts oriented. Art activities are quite varied; breadth is emphasized over depth. The art instructors have ample supplies and, although the building is old, the facilities are adequate and the art room is ample in size.

The subjects

It is questionable whether any group or class of art students would be a representative sample of Wilson Junior High School students. Although the art classes are not grouped by any selective criteria, other subjects are, and it is possible such grouping would reflect in the make-up of any one art section. A more likely source of bias has to do with the selection of electives, especially the choice between art and band or choir.

Band is more in keeping with middle class values. It is a "clean" activity and is consistent with the American core value of the "group common man." As a performing art, band has extrinsic rewards for both parent and child, i.e., public recognition and the opportunity to travel. Economics, in part, pushes many children of lower economic class parents into art, and some of these children are attracted to art because of its relationship to hand skills. Art classes as a "dumping ground" for underachievers and non-college bound students still persist--there is no place else to put some children.

Two basic considerations, in sum, stipulate that the subjects chosen for the study would have a negligible background of formal art instruction, but at the time of the study they would be near the earliest level of the education in which their art instruction was being conducted by a qualified, state certified art instructor. Ordinarily, this would mean the seventh grade; however, an eighth grade art class was chosen for two reasons. First of all, upon the advice of the art instructors involved, the eighth grade class was selected because there was less likelihood that art students at this level would change courses at mid-year. Secondly, the eighth grade art class selected for this study was under the tutelage of two art instructors. This "team teaching" situation offered the opportunity of having two persons who would be in intimate daily contact with the subjects. Thus, more reliable observations could be gathered about the subjects.

The study began in December, 1965, with 24 subjects. At the end of January, 1966, one more subject was added to the original number as there appeared to be a danger of depletion by the transfer, course changes, and irregular attendance by a number of the subjects. The study was completed in late May, 1966. At this time there were 22 subjects who had completed at least one-half of the art assignments during the period of the study.

An effort was made in the early stages of the study to encourage all subjects to complete all assigned art projects, but such attempts were discontinued when it was found that the laggards became further delinquent in succeeding assignments. Thenceforth, only those subjects who were near completion at the end of an art activity were encouraged to complete that particular task.

Art Projects and Presentation

Art projects

It had been predicted at the outset of the study that approximately eight to ten art projects would be completed during the five month period of the study. Actually, nine art projects were completed; however, one was deleted. The deleted project was illustrative materials done in conjunction with a social studies project. Size, descriptive nature, and questionable sources of illustrations were the reasons for its deletion. The projects and brief description follows:

Two-dimensional:

1. Cartoon on school life (crayon)
2. Prints of pictorial and pure designs (linoleum block prints).
3. Ceramic designs (crayon)
4. Designs of naturalistic subjects applicable for papier mâché plaques (crayon)
5. Fantasy drawings of animal types (crayon).

Three-dimensional:

1. Ceramics (glazed pottery)
2. Ceramic sculpture (glazed)
3. Papier mâché plaques (tempera finish)

Art products presentation

For ease of presentation to the judges, it was decided to photograph the art products and to present each art task as a unit within a single format. Photographs of art works as visual aids in art analysis are not new, but in research designs they are relatively uncommon. Beittel and Burkhart used photographs to record procedural orientations of their subjects while they were performing art tasks.[4] Besides the ease of collecting and transporting art products to judging sites, there were three reasons which lent support to the use of

---

[4]Kenneth R. Beittel and R. C. Burkhart, "Strategies of Spontaneous, Divergent, and Academic Art Students," Studies in Art Education, 5: 20-41, No. 1, (1963), p. 21.

photographs. First of all, photography allowed all art projects to be seen in the same "light." Secondly, the format of a photograph aids the concentration of vision. Finally, the presentation of all art products from a single art task within a single format made ranking procedures and comparisons easier.

It may be questioned whether the photographing of art products does not lead to the loss of some of the more subtle qualities of the art work. The experience of photographing and analyzing the pre-study junior high art products which were used for the purpose of training art judges did not lend any evidence to this possibility. For one thing, a junior high-aged art student generally has not achieved a level of sophistication in handling art materials which would make subtle qualities a determining factor in the ranking of art products. In fact, a photograph of an art product tends to enhance its over-all quality; the photograph is more visually pleasing than the real item. It is assumed that this is an equally effective characteristic of all photographed art products. Furthermore, it is unlikely that any of the elements of the two criterion measures used in this study would be materially influenced as a result of photographing the art products.

The format for the presentation of the photographed products of each art task was a 22" x 28" white mat board. There were 24 mounts for photographs on the board with each mount having a hinged flap. The flap in the closed position (covering mount and photograph)

revealed an identifying letter. Each letter was given an ordinal rank during the judging procedure. All identifying marks were eliminated from the face of photographs, but each photograph had a number on its back for student identification purposes.

Ranking procedure

"Forced" ranking was used by the judges; that is, ordinal ranks were alternately assigned to both ends of the ranking continuum. In this proceduring, rank number one would be assigned to that art product which was considered "best" in terms of the defined elements of the criterion used in judging; the lowest ordinal rank was then assigned to that art product which was considered "worst" by an art judge in terms of the same defined criterion elements; then, the "next best" was chosen, the "next worst," and so on until all of the art products were assigned an ordinal rank. In this way the less differentiated middle group of art products would be reduced in number for easier rank assignments.

The Teacher Judges ranked the subjects, rather than art products, on the termination of the study in terms of both criterion measures. This was necessitated by the teacher's intimate knowledge of the students and their art products.

Statistical Analysis

Nonparametric statistical tests

The choice of an appropriate statistical test for the analysis

of data is unquestionably important. Some of the reasons for choosing nonparametric statistical tests for this study are obviously apparent; however, there are less obvious reasons which not only determined the choice of statistical test but also determined the structure of this research design.

One of the more obvious reasons for relying on nonparametric statistical tests is that these tests focus on the order or ranking of scores, not on their "numerical" values which makes judgments easier to make. That is, a "more than/less than" decision would be easier for judges to make than trying to determine exact "scores" for art products which by their nature do not lend themselves to exact measurement.

The less obvious, and more important, reasons for choosing nonparametric statistical tests, which were also determinants of the research design, have to do with the reluctance of accepting two basic assumptions which underly the parametric tests, especially the 't' test. First, one of the conditions which must be satisfied to make the 't' test the most powerful one, and in fact, before any confidence can be placed in any probability statement obtained by the 't' test, is that the observations must be drawn from a normally distributed population. Although an attempt was made to obtain normalcy in population distribution by selecting a population which displayed such characteristics, it became apparent, due to the stated, underlying

limitations of this study, that the attempt was less than successful.[5] For this reason, "distribution free" or nonparametric techniques were chosen as they result in conclusions which require fewer qualifications, and probability statements obtained from the selected nonparametric statistical tests are exact probabilities, regardless of the shape of the population and sample distribution.

Secondly, a most important determiner of the structure of this research design and choice of statistical model had to do with the two criteria being measured in this study. An underlying assumption of the 't' test is that the variable involved must be measured in at least an interval scale so that it is possible to use the operations of arithmetic (summing, finding means, standard deviations, etc.). The authors of the two criteria used in this study assigned numerical scores of an interval nature to their observations. They were willing to assume that the structure of their measurement scales was isomorphic to that numerical structure known as arithmetic; that is, they assumed that they had attained a high level of measurement. It is highly questionable that they had attained such a high level of measurement that their scoring was amenable to an interval scale and the manipulations of arithmetic. It is highly questionable, for instance, in measuring the spontaneity criterion' element "media overlap" that an

---

[5]The problems of selecting a population for this research have been discussed in the first part of this chapter.

art product of subject "A" displays three times as much of this element as does an art product of subject "B," or this same art product of subject "A" displays three times as much creativeness through "novel elaboration" as does the art product of subject "B." In essence, this is what a judge is saying when he assigns a score of 3, in terms of these stated criterion elements, to the art product of subject "A" and a score of 1 to the art product of subject "B." The point is, although the scores appear more precise than ranks, these scales do not meet the requirements of any higher level of measurement and may properly be viewed as ordinal. Sidney Siegal has this to say about the problem of measurement and statistical analysis:

> ---the writer wishes to emphasize here that parametric statistical tests, which use means and standard deviations (i.e., which require the operations of arithmetic on original scores), ought not be used with data in an ordinal scale. The properties of an ordinal scale are not isomorphic to the numerical system known as arithmetic. When only the rank order of scores is known, means and standard deviations found on the scores themselves are in error to the extent that the successive intervals (distances between classes) on the scale are not equal. When parametric techniques of statistical inference are used with such data, any decisions about hypotheses are doubtful probability statements derived from the application of parametric statistical tests to ordinal data are in error to the extent that the structure of the method of collecting the data is not isomorphic to arithmetic.
>
> In as much as most of the measurements made by behavioral scientists culminate in ordinal scales (this seems to be the case except in the field of psychophysics, and possibly in the use of a few carefully standardized tests). This point deserves strong emphasis.[6]

---

[6]Sidney Siegal, Nonparametric Statistics for the Behavioral Sciences, New York: McGraw-Hill Book Company, Inc., 1956, p. 26.

The willingness of an investigator to accept or reject the underlying assumptions of parametric tests is not a clear-cut, pre-defined decision, for as Siegal points out:

> Although some empirical evidence has been gathered to show that slight deviations in meeting the assumptions underlying parametric tests may not have radical effects on the obtained probability figure, there is as yet no general agreement as to what constitutes a "slight" deviation.[7]

From the level of measurement achieved by the two criterion scales used in this study and the lack of standardization of these scales, it is posited that there is much more than a "slight" deviation from meeting the assumptions underlying parametric tests.

The only assumption made by the nonparametric statistical tests used in this study (an assumption which also underlies parametric tests) is that the observed scores are drawn from an underlying continuous distribution of the variate under consideration. This assumption is acceptable, and, as it was pointed out earlier, an attempt was made to assure a wide spread of the criterion variates (elements) to be measured in this study by selecting a hetereogeneous population for the study.

Finally, the modification of the criterion scales used in this study from an interval scale to an ordinal scale is an acceptable practice; that is, a transformation which does not change the order of classes is completely admissable because it does not involve any loss of information.

---

[7] Ibid., p. 20.

Two nonparametric statistical tests were used for data analysis in this study. They were 1) Kendall Coefficient of Concordance: W,[8] and 2) Spearman Rank Correlation Coefficient: r_s.[9] The rationale for the use of each of these statistics will be stated in conjunction with their application for a particular purpose.

---

[8] Ibid., pp. 229-238.

[9] Ibid., pp. 202-213.

CHAPTER III

JUDGES AND JUDGING

A major part of this study required the selection and training of personnel to serve as judges for evaluating art products. High interjudge correlation in the use of the selected criterion measures was necessary if any degree of confidence could be assigned to implications drawn from the acceptance or rejection of the stated hypotheses. It is for this reason that a separate chapter is devoted to this phase of the study. This chapter will deal with 1) selection of judges, 2) training of judges, 3) interjudge correlations within each judge group, and 4) observations of the results of this ongoing process.

Selection of the Judges

There has been a consistent trend in research designs that have incorporated the use of judges to use personnel who were considered well-versed in the area being investigated or considered "expert" in the discipline in which the investigation was taking place. The choice of judges for this study was consistent with this trend. All judges were either teaching or had taught art and had an advanced degree or were working toward an advanced degree in Art Education.

The Typology Judges, who were to evaluate art products on the basis of the criterion measure, "Typology of Creativity," were three

teachers of art at Roseburg High School in Roseburg, Oregon, which lies 75 miles south of Eugene. The Six Criterion Judges, who were to evaluate art products on the basis of the criterion measure, "Six Criterion Elements of Spontaneity," were three graduate students in the Department of Art Education at the University of Oregon, Eugene.

The Teacher Judges, who were to use both criterion measures to rank subjects rather then art products, were two teachers who were responsible for the program and the teaching of art at Wilson Junior High School in Eugene.

The separation of judge groups in terms of geographical distance eliminated any possibility of intra-judge group communication on their evaluations. Furthermore, it became apparent during the training sessions that the judges were not familiar to any degree with the criterion measures, although two of the judges knew of the better known measure of spontaneity. Both of these judges belonged to the Six Criterion Judge group who was using this particular criterion for evaluating art products. The possibility that a judge or judges would investigate the criterion measure of another judge group could not be eliminated; however, there was no indication from the judges that this had happened. Furthermore, the weight of the judges' regular obligations and the shortness of time devoted to the training in the use of the criterion measures plus the rapidity of receiving sets of art products for evaluation made such possible contamination remote.

Kendall's Coefficient of Concordance: W

For the Six Criterion Judges and the Typology Judges, Kendall's Coefficient of Concordance: W was used to determine interjudge agreement in the ranking of art products. The choice of this statistic is best explained by Siegal:[1]

> When we have 'k' sets of rankings, we may determine the association among them by using Kendall's Coefficient of Concordance: W. Whereas 'rs' and 't' express the degree of association between two variables, W expresses the degree of association among 'k' such variables. (p. 229)

'W' bears a linear relation to the average 'rs' taken over all groups. If we denote the average value of the Spearman rank Correlation coefficients between $\binom{k}{2}$ possible pairs of rankings as 'rs_{av}' then it can be shown that

$$rs_{av} = \frac{kW - 1}{k - 1}$$

Training of Judges

The initiating training sessions for the Six Criterion Judges and the Typology Judges were conducted near the end of the study (mid-April) and consisted of 1) defining and discussing the two criterion measures, and 2) evaluating, as a group, examples of art products which were at the near-grade level of the subjects' art products under study.

---

[1]Sidney Siegal, Nonparametric Statistics for the Behavioral Sciences, New York: McGraw-Hill Book Co., Inc., 1956.

After this initial session and within the next two weeks each of the judge groups was presented five sets of photographed art products to judge for training purposes. Both judge groups were presented the same sets of art products.

The art products which composed these training sets were done by subjects who were at the grade level or near-grade level of the subjects under study. Each judge made his evaluations independently of all other judges. The results of this training period, including the art tasks and the number of art products judged, are summarized in Table I. These correlations seemed sufficiently high to satisfy the requirements for interjudge reliability.

Although one cannot draw conclusions or make definitive statements from such meagre data as summarized in Table I, the results are suggestive in certain respects.

The most apparent question which could be asked is, "Why is there such a low correlation between judges within both judge groups on the evaluation of art products of the fourth art task, non-objective paintings?" From discussions with the judges, there was agreement amongst them that considerable difficulty was experienced in evaluating the products of this art task. Possible reasons for this difficulty can be forwarded from a comparison between a description of the process and results of this art task and the defining statements of the two criterion measures. The objective of this art task was to offer the students an exercise which would demand their becoming involved in

TABLE I

CORRELATION COEFFICIENTS (W) OBTAINED WITHIN JUDGE GROUPS ON TRAINING TASKS[2]

Order of art tasks presented	Number of art products	Typology Judges	Six Criterion Judges
1. Papier mâché masks	10	.785***	.930**
2. Rock mosaic	10	.801***	.639****
3. Pencil drawing (nature)	15	.806**	.933*
4. Non-objective paintings (ink and wax resist)	18	.324	.369
5. Watercolors (nature)	15	.796**	.887*

*significant at the .001 level of confidence
**significant at the .01 level of confidence
*** significant at the .02 level of confidence
****significant at the .05 level of confidence

[2]All positive correlations. It could not be otherwise as one difference between 'W' and the 'rs_{av}' methods of expressing agreement among 'k' rankings is that 'rs_{av}' may take values between -1 and +1, whereas 'W' may take values only between 0 and +1. The reason that 'W' cannot be negative is that when more than two sets of marks are involved, the rankings cannot all disagree completely. That is, when more than two judges are involved, agreement and disagreement are not symmetrical opposites. 'k' judges may all agree, but they cannot all disagree completely. Therefore 'W' must be zero or positive.

art process rather than being chained to a preconceived image of the final art product. Each step of the process had to be completed before the nature of the succeeding steps could be determined. The media used were waterproof colored inks and wax. The ink is transparent which is conducive to overlays, and the wax resists the ink and thus renders chosen areas of the painting impervious to color. Although a few non-objective shapes were drawn on the surface at the beginning, it was quite possible these shapes would be completely lost from view in the final product. That is, after the first application of wax and a color were applied and before the next application of wax and color were applied the work was studied for potentially new, emerging shapes and color relationships. In most cases, the final results were quite different from whatever planning had taken place in the initial stage of the process, and with the use of the wax medium, which is not conducive to precision of line or shape, it gave the final products a look of spontaneity which may not have emerged otherwise.

Eisner's "Typology of Creativeness" identifies 'boundary breaking' as the highest type of inventiveness, and the locus 'form' under this category is defined as "the formulation of utterly new forms." It is quite possible that all of the Typology Judges saw all products of this art task as meeting this definition. If this were the case, how does one determine to what degree some art products are more "utterly new" than others?

The Six Criterion Judges could have experienced difficulty in their evaluations, because the "Six Criterion Elements of Spontaneity" were quite apparent and of high value in the products of this art task. That is, the art products displayed the spontaneous characteristics of moving rather than still objects, movement within shapes, loose detail, media overlap, ragged border of shapes, and the lack of mechanically imposed texture. The Six Criterion Judges may have achieved a higher interjudge correlation than the Typology Judges because the six elements used for evaluating spontaneity are not as restrictive as trying to rank order "utterly new forms."

If these arguments are valid and if this were the case during the evaluation of this training set, then there are possible detrimental implications in the use of these criterion measures. The most obvious is the difficulty judges could experience in evaluating products of certain art tasks. Another point which can be made is that the system can be fooled. Suppose, for example, that children were being selected from several schools for a class of students gifted in art, and this giftedness were to be ascertained by their creativeness in art work as determined by these two criterion measures. In this situation, a teacher could have a greater representation from his art classes than that of other participating teachers by carefully presenting art tasks which required novel solutions in terms of subject matter and form and by limiting the media used in the execution of these art tasks to those which are not conducive to

precision of line and shape. Even in the normal situation, different teachers will vary in their choice of art tasks and media, so even without intent, selection of gifted children may not be accurate or just.

Since the training sessions were conducted well toward the end of the collecting of art products of the subjects involved, and the nature of the concluding art tasks was known, the possibility of this kind of evaluation difficulty was remote.

The return to a high correlation within judge groups on training set number five indicated there was no change in the judges' confidence in using the criterion measures because of their experience with the non-objective paintings.

Further perusal of Table I shows that the Six Criterion Judges were consistently higher than the Typology Judges in interjudge correlations, except in one instance. The products of the rock mosaic art task had somewhat the same visual qualities as the non-objective group. Made of pea-sized, colored rock and without the usual precise, dark colored contour line around the shapes which is often identified with mosaics, the final results were often characterized by the defining high value attributes of the criterion spontaneity. Much of the loose detail and ragged contour lines was due to the artists' lack of skill. For the Six Criterion Judges this probably made the evaluation of the rock mosaic products difficult because nearly all products had a high degree of spontaneous-like qualities. On the

other hand, the Typology Judges would not necessarily have had the same kind of difficulty in evaluation. The subject matter of the mosaics was derived from naturalistic forms and the delineated shapes (forms) adhered to the general concept of the naturalistic forms; i.e., there was little innovation of form. Finally, it was the consensus of opinion of the Typology Judges that rank-ordering in terms of innovation of subject matter was easier than determining the rank-order in terms of innovative form.

Although the result of the training period did not indicate a measurable growth of confidence on the part of the judges in using the criterion measures, there were consistent correlations within judge groups across art tasks, and they seemed sufficiently high to satisfy the requirement for interjudge reliability. Furthermore, any statistically measurable growth of confidence on the part of the judges in using the criterion measures would have to be both consistent and fairly dramatic before it could be said that the growth was not due to the nature of the art tasks being judged or to the accidental presentation of the art task products in the rank-order of ease of judgment.

Finally, it was pointed out in Chapter I in discussing the employment of art juries for evaluating art products, the statistical danger in pooling (averaging) judgments, i.e., an art judge who is at odds with the other judges of an art jury may bias the "true" rank value of an art work. This investigator is aware of this possible

source of bias but is confident that increased precision of pooling technique is unwarranted in view of other imprecisions in the data collected in this study.

Interjudge Correlation in the Evaluation of the products of Eight Art Tasks

Table II shows that moderate to very high interjudge correlations were achieved across the eight art tasks. The difficulty encountered during the training session with the evaluation of non-objective paintings does not have a comparable experience during the judging of art products of these art tasks.

In consideration of the two judge groups and the two criterion measures, the interjudge correlations obtained on the art products of the tasks and the art materials and processes used in executing these tasks can be justified.

Cartoons, because of their simplified symbology and the lack of any inherently spontaneous qualities of the medium (crayon) used, were simple in concept, yet differentiated enough in subject and surface treatment to make rank ordering fairly easy. The lower interjudge correlation obtained from the Typology Judges could be a result of nearly all art products being classified as being of one type of creativity (the lowest) and mainly within one locus (subject).

One generally thinks of prints as being quite sharp and clear (non-spontaneous appearing), however, the surface results of many of the prints were characterized by loose and randomly placed textural

TABLE II

CORRELATION COEFFICIENTS (W) OBTAINED WITHIN JUDGE GROUPS ON THE EIGHT ART TASKS

	Order of Art tasks presented	Number of art products	Typology Judges	Six Criterion Judges
1.	Cartoons (crayon)	22	.682**	.759*
2.	Prints (waterbase inks)	21	.868*	.790*
3.	Ceramic designs (crayon)	20	.754**	.804*
4.	Ceramics (pottery)	16	.643***	.899*
5.	Ceramic sculpture	17	.931*	.794**
6.	Papier mâché designs (crayon)	16	.633***	.898*
7.	Papier mâché plaques (tempera)	19	.569****	.664**
8.	Fantasy drawings (crayon)	16	.876*	.887*
			Ws_{av} .743**	Ws_{av} .812*

*Significant at the .001 level of confidence
**Significant at the .01 level of confidence
***Significant at the .02 level of confidence
****Significant at the .05 level of confidence

effects. This was probably due to not enough ink, lack of uniform pressure, and/or the ink drying too quickly. These prints could serve as examples in which the lack of skill in handling the medium rather than the personality of the artist results in spontaneous-like effects. The very high interjudge correlation obtained by the Typology Judges was somewhat of a surprise because the prints were of both non-objective and naturalistic forms, and the judges had experienced difficulty when trying to determine which of the two loci, subject or form, is of higher value when both are at the same level (type) of creativity and display the same degree of the defining characteristics. One explanation is that many of the non-objective prints were composed of simple, familiar geometric forms (shapes) and may not have been considered very innovative. The prints task was one of two tasks in which the Typology Judges surpassed the Six Criterion Judges on the obtained interjudge correlation coefficient.

The ceramic designs which were drawn on paper with crayon were similar to the cartoons in result. They were simple, yet differentiated designs, and of one type of creativity (lowest) but within a different locus (form) than the cartoon products. As with cartoons, similar observations could be drawn from the products of ceramic design.

The products of the ceramics task, which were comprised of the usual pots, ash trays, and candlestick holders, showed a wide range of surface treatment, but again, the degree of spontaneity in handling the

material was probably more attributable to the artist's lack of skill than to his personality. The lower interjudge correlation obtained by the Typology Judges could be attributed, in part, to the general low level of creativeness displayed by the art products and the similarity of forms between several of the examples.

The ceramic sculpture task is the only other instance in which the interjudge correlation of the Typology Judges was higher than that of the Six Criterion Judges. Observations about the art products of this task along with the ceramic task will be dealt with more fully later in the chapter. However, an observation is in order at this time in regard to the art task and the Six Criterion Judges. Invariably, the sculpture products of an animal form were quite commonplace and rendered, in most instances, in a stiff, mechanical-like way. These results could have been caused by unfamiliarity with the animal and/or medium, and the lack of looseness in handling the medium probably made rank-order judgments difficult.

The papier mâché designs were very similar in concept to the art products of the cartoons and ceramic designs, and the interjudge correlations achieved by the two judge groups were also similar.

Table II shows that the papier mâché plaques, which were an outgrowth of the preceding art task, resulted in the lowest interjudge correlations for both judge groups. Observations from the results of this kind of 'before and after' activities will be dealt with more fully in Chapter IV which deals with the testing and

implications of the stated hypotheses of this study. In focusing on the products of this art task, however, the products were considered, in general, the least interesting and original of all the art products judged. For the subjects, papier mâché was not a satisfactory art experience. The problem was that the subjects lacked the foresight in their planning as to the limitation and potential of the material they would use in the execution of their designs; the degree of detail exhibited in the art products of the papier mâché design was not conducive to the bulky quality of papier mâché. The modifications required of the original designs and the loss of interest in the art task due to dissatisfaction resulted in unimaginative projects whose surface treatments were more crude and indistinguishable rather than free and spontaneous.

The fantasy drawings exhibited quite a range of unusual forms and subjects, and due to the unnaturalistic orientation of the art task, the subjects were quite free in the surface treatment of their drawings. Probably both judge groups found these art products easy to rank order, and in any case, both judge groups achieved very high interjudge correlations.

Table II shows that the Six Criterion Judges were more consistent and generally higher in interjudge correlations in the judging of the products of the eight art tasks than the Typology Judges.[3] Except for

---

[3] 'Ws_{av}' achieved by the Six Criterion Judges was .812 which reached the .001 level of significance, and the 'Ws_{av}' achieved by the Typology Judges was .743 which was significant at the .01 level.

consistency of correlations, this was somewhat true of the results from the training session. Possible reasons for this difference of inter-judge correlations between judge groups could have been due to some physical or personality factors of the judges who composed the judge groups. Age, sex, role identification, environment, personality clashes, repugnance of task, lack of or differences in understanding of the defining concepts of the criterion measures could have been influencing variables in the ability of a judge group to achieve and maintain high interjudge correlations; however, it is in the analysis of the Typology Judges that possible reasons for intra-judge group differences may exist.

Knowledge and observation of the judges and the statistical analysis of judge pairs within the Typology Judges suggest art training, nature of art tasks, the defining statements of the criterion measure, and the rank order method of using the criterion measure were possibly greater influences than any physical or personality factors in determining the level and consistency of interjudge correlations of this judge group.

The Typology Judges, like the Six Criterion Judges, were young, and their level of education and amount of teaching experience were similar. There is no apparent source for variation due to these judge characteristics. It hardly seems likely that sex was a variable as both judge groups were composed of both sexes; Typology Judges were composed of two male judges and one female; whereas, the Six Criterion Judges

had the reverse proportions. Personality clashes must be ruled out as the judges evaluated the art products independently, and there certainly was no conflict between the individual judges and this investigator. The judges were well known to this investigator, and there was the utmost confidence in the sincerity of interest expressed by the judges in this study. The judges were diligent and prompt in performing the evaluation tasks.

An analysis of paired correlations of judges in this judge group in relation to some comments made by the judges in post-evaluation discussion appear to offer more promising leads for explaining the lower and less consistent interjudge correlations of the Typology Judges across art tasks.

Table III, in which the paired correlations of the Typology Judges are summarized, shows that no judge was consistently at odds with the other judges in this judge group.[4] Judge A3 (male) had the lowest correlation with the other two judges on art tasks 3, 6, 7, and 8, but otherwise low paired correlations were evenly distributed with each of the other judges being lowest on two tasks

Although there is no clear pattern of deviation among the Typology Judges, a comment by one of the judges suggests the possibility that a judge's background and interest in art may influence his perceptual

---

[4]The computation of paired correlations within the judge group required reverting to the more familiar statistic, Spearman Rank Correlation Coefficient: rs.

TABLE III

PAIRED CORRELATION COEFFICIENTS (rs) OF THE TYPOLOGY JUDGES OVER THE EIGHT SETS OF ART PRODUCTS

Judge Pairs	Art Tasks							
	1	2	3	4	5	6	7	8
A1 - A2	.393	.756	.810*	.429	.811	.514*	.489*	.844*
A1 - A3	.878*	.829*	.695	.247	.818	.317	.304	.797
A2 - A3	.583	.766	.738	.718*	.939*	.485	.286	.813

*highest paired correlation within each art task.

ordering when looking at art work. Judge A2, whose interest and art training are in sculpture, commented that he tended to view the art products in terms of form rather than the other art elements (line, texture, color).[5] Table III shows that this judge (A2) had highly significant correlations with judge A3, who was responsible for teaching pottery and ceramic sculpture, on the only two art tasks with products which were clearly three-dimensional in concept. These were the only two instances in which these two judges' paired correlations were higher than the correlation pairings with the third judge. No other judge in either judge group had the training or immediate involvement and interest in three-dimensional art work at the time of this study.

Although it is conjecture, this writer strongly suspects that one's interest and involvement in a particular art form will influence that person's perception of art works. If this is so, then the selection of judges for the evaluation of art works must be more

---

[5]Form is a term which has many definitions and leads to misunderstanding even within such a specific area as art. Judge A2 was referring to three-dimensional form and the emerging and receding relationships between forms; apparently, this way of perceiving was projected into the two-dimensional works of art being judged. Form, as defined in the criterion measure, "Typology of Creativity," is more comprehensive and would include two-dimensional shapes as well as three-dimensional forms. Form, by this definition, would also include flatly executed areas of color which would be seen by some artists as neither shape or form.

restrictive in qualifications than a general expertise in art. Not only would the nature of the judges' training and interest in art and the kind of art form from which the art products were derived be factors to consider in judge selection, but also the definitions of the criteria measure used in the evaluation process would be a consideration. For instance, Eisner gave as an example of boundary-breaking creativity to be a student who makes his painting upon a sphere rather than a flat surface.[6] The question is whether a sculptor, with his experience with decorating and enhancing a multitude of various topographical surfaces, would perceive Eisner's example as being as novel as would a painter who has a more stereotyped perception of two-dimensional surfaces?

Another possible reason for the Typology Judges having lower and less consistent interjudge correlations, and one which this investigator suspects as having the greatest influence, was the use of rank ordering as a method for measuring the degree of the defining characteristics of the "Typology of Creativity" displayed in art products.

During the post-evaluation discussion, one of the judges said, in retrospect, that in cases in which it was difficult to determine the rank order of two or more art products he may have relied on the degree of complexity rather than novelty for his selection. The other

---

[6]Elliot W. Eisner, "A Typology of Creativity in the Visual Arts," Studies in Art Education, 4:11-22, No. 2 (1962), p. 12.

two judges were prone to agree with this observation. Apparently, this change of judging criteria was done unconsciously. The possibility that the judges shifted to other criteria in judging art products is real enough to warrant reviewing the rationale and validity of using the rank order method evaluation at this time.

Eisner's types of creativity are, in a sense, on a continuum in which the degree of novelty exhibited by art products determines their place on this continuum and their type of creativity. Each of Eisner's judges evaluated the art products on a nine-point scale and used a separate criterion element for each of the two loci (form and subject) subsumed under each type of creativity. The judges had to make two types of decisions. "They decided (a) if the characteristics they sought were displayed and (b), if the characteristics were displayed, to what extent they were evident."[7]

It is clear that one of the underlying assumptions of this criterion measure is that one or more of the characteristics being evaluated may not be present in the art products, but what is not clear is whether none of these characteristics is present.

This investigator was willing to accept the assumption that the variable under study was normally distributed along a continuum, but an unacceptable assumption was that the measurement scale used in the Eisner study (nine-point scale) was so precise as to make it

---

[7] Ibid., p. 14.

amenable to an interval scale and the manipulations of arithmetic. In part, it was for this reason that the rank order method was chosen as the measurement scale for this study.

Furthermore, the assumption was made by this investigator that at least one of the characteristics identified in Eisner's "Typology of Creativity" would always be present to some degree in the art products evaluated. There were two reasons for making this assumption: 1) the general emphasis of qualified art teachers on originality of art results, and 2) the naivete and lack of art skills of the young artists would produce art products which would appear, to some degree, unique and novel. This assumption is now in question because the judges apparently evaluated some art products as possessing none of the defining characteristics of their criterion measure. For example, in reviewing the art products in this study, it was noted that there were instances in which two or more subjects had chosen the same subject and interpreted the subject in a similar fashion. For instance, two young artists had chosen an alligator for their subject in the ceramic sculpture task. Probably due to a lack of a sophisticated concept of an alligator, the final products seemed to indicate that the artists' energies were solely devoted to achieving a naturalistic interpretation of their subject. The results were similar and the differences were in the size and the amount of detail. It could still be argued that the art products of these young artists were "novel" because of their naive and unskillful interpretations. However,

it seems clear that the intent of the artists was a naturalistic interpretation and the attempt to be original was nil. The outcomes were sterile.

The conclusion drawn from this analysis is that the Typology Judges, when faced with two or more similar art products, used criteria alien to the "Typology of Creativity" to determine the rank-order of the art products, and it would appear that, unconsciously, the judges made their choices on the basis of the degree of complexity exhibited in the art products. It is likely that this dilemma would not be faced if a raw score were assigned to each of the characteristics of the criterion measure, "Typology of Creativity," which was displayed in a particular art product. If this conclusion is valid, then an investigator, who assumes that a variable under study is not amenable to an interval scoring scale or the operations of arithmetic and that there exists the possibility that the variable under study may have a value of zero in some of the material evaluated, must choose between the lesser of two evils or select or devise a means of measurement which would eliminate the need for such a decision.

There are at least two strategies which can be employed to insure a greater degree of consistency of interjudge correlations throughout the evaluation period. Other researchers who have used judges in their studies have employed these strategies 1) to guard against the fatigue factor or changes in the application of criteria during the evaluation period and 2) to reduce extreme ranking or rating

differences between judges on the evaluation of particular products. One of the strategies used is the employment of a "third person" to spot-check the reliability of individual judges in the employment of the criteria; the second strategy is to discuss and re-evaluate products in which the scoring or ranking resulted in wide deviation among two or more judges. It was the decision of this investigator to employ neither strategy.

The strategy of using a "third person" to spot-check the reliability of the on-going evaluations of the individual judges was discarded because it was felt that neither the number of art product sets nor the length of the evaluation period (Including the training period, there were 13 sets of art products and an over-all evaluation period of five weeks) was so great as to influence the outcome of the evaluations.

The second strategy, that of discussing and re-evaluating art products which had caused wide deviation of scores or ranks among judges, was rejected because of its potential source of bias in the judging procedure. It was felt that the judges may be too easily swayed from possible justifiable evaluations due to either an over-willingness to please the investigator or the acceptance of him as an all-knowing authority figure.

In this analysis of the Typology judges, one must not lose sight of the Six Criterion Judges who had, on the most part, very high and

consistent interjudge correlations. It is clear from the comparison of these two judge groups that the criterion measure used by the Six Criterion Judges was simpler to use, and the possible reasons for variance of interjudge correlations within the Typology Judges apparently did not influence their evaluations. The global criterion, spontaneity, requires a judge to make, basically, one decision, i.e., the degree of spontaneity an artist has exhibited in handling the art materials when treating a surface. The six elements of spontaneity are focal orientations for a judge to employ in scanning the surface of an art product. Art training and experience and, in most cases, the nature of the art task have no relevance to this kind of decision-making; it makes the more than/less than decisions of the rank-order method of measurement quite simple. Furthermore, this criterion measure has a built-in complexity factor, although it is the looseness in handling detail rather than the amount of detail which determines decision-making.

In addition, Eisner's "Typology of Creativity" measures one criterion: creativity. However, the defining characteristics of each type of creativity and each of the two loci which comprise each type of creativity are quite distinct and do not lend themselves easily to a composite evaluation. For example, two art products may be judged to be at the same level of creativity, one on the basis of the locus, form, and the other on the locus, subject. There are no criteria for determining which of the loci would take precedence over the other. This is

another example of an instance when a judge may rely on some criteria alien to the criterion measure to determine rank-order of art products.

Teacher Judges

One of the attractive features for selecting the group of subjects used in this study was that there were two teachers involved in the teaching of art to the subjects. This was an opportunity to observe how well a judge team who was intimately involved with the subjects would rank their students on the basis of the two criterion measures and to compare their evaluations with the results of the other two judge groups. Due to the intimate relationship between the teachers and the subjects, the teachers were asked to rank the subjects rather than their art products upon the termination of the study. The reason for comparing Teacher Judges with the other judge groups was to see whether the teachers' knowledge of the subjects and the observation of the subjects' behavior during the process of executing art works would result in a significant difference between the ranking of subjects by the Teacher Judges and the ranking of subjects by the other two judge groups. The ranking of subjects by the Typology Judges and the Six Criterion Judges was to be determined by pooling the rankings of the subjects' art products by each of the two judge groups. If differences did exist, there was to be a follow-up discussion with the teachers to determine what factors might have been influential in the manifestation of ranking differences. If, in fact,

there were no significant ranking differences between the three judge groups, then there would be evidence that factors not evident to the Typology Judges and the Six Criterion Judges were of little relevance in the employment of the criterion measures and that the Teacher Judges' evaluation of the subjects would compare favorably with judges who were not involved with the subjects and were making their evaluations on the basis of the subjects' art products.

However, the results of interjudge correlations between the two teachers, as summarized in Table IV, cast doubt on the value of following through with the planned procedure.[8] It is quite evident that the two judges saw their subjects in an entirely different light.

TABLE IV

'rs' VALUES OBTAINED BETWEEN THE TEACHER JUDGES IN RANKING THE SUBJECTS

Criterion Measure	
Six Criterion Elements of Spontaneity	Typology of Creativity
-.449**	-.580*

*significantly negative correlation at the .01 level.
**significantly negative correlation at the .05 level

[8]In order to determine interjudge reliability, Spearman Rank Correlation Coefficients: rs were computed from the ranks.

Both judges were very cooperative during the collection of art samples and the evaluation period, but there were observable signs of friction between them. Except for hearsay, there was little evidence as to the fundamental causes for these differences, and as the pursuance of the reasons for these differences offered little in value for the purposes of this study, the matter was not pursued.

Due to the results of evaluations by the Teacher Judges, the research design had to be modified, and this judge group will not play a significant part in this study.

Summary

This chapter was devoted to describing the selection, training, and the performance of the judges employed in this study, the analysis of obtained interjudge correlations from their evaluations of the subjects and their art products, and suggestive results from the analysis of this part of the study.

The Typology and Six Criterion Judges had little difficulty in grasping and effectively employing the defining characteristics of their respective criterion measures. Interjudge correlation coefficients from the outset of the training period were from moderate to very highly significant. However, the Typology Judges generally had correlations lower than the Six Criterion Judges throughout the judging period.

Although there was no conclusive evidence obtained from the analysis of the judging experience, there were some suggestive results. The observations made from these results should be given due consideration in future studies which incorporate judges in their procedure, or they could serve as a basis of further experimental research designs. The following tentative conclusions were made.

1. The criterion measure, "Six Criterion Elements of Spontaneity," can generally be reliably employed by judges employing the rank-order method of evaluating art products, because only one "more than/less than" decision needs to be made. This decision, which is not easily influenced by other variables, is based on the treatment of the surface of the art product.

2. Art tasks which employ art materials or processes which inherently result in a free, and "spontaneous-like" appearing quality can cause lower interjudge reliability in the employment of the criterion measure, "Six Criterion Elements of Spontaneity." Furthermore, such art task products could bias the selection of students from multiple classes or schools for special art classes if this criterion measure served as a basis for such selection.

3. The criterion measure, "Typology of Creativity," is comprised of distinct defining characteristics which, at times, do not lend themselves to the composite judgment necessary in the rank order method of evaluation. When two or more art products display different, but equal in value, defining characteristics, a judge will rely on criteria alien to this criterion measure in determining the rank-order of the art product.

4. A judge's art training and experience may influence his perceptual ordering and emphasis of the defining characteristics of the criterion measure, "Typology of Creativity."

5. The nature of art tasks, such as non-objective art work, can cause lower interjudge reliability in the employment of the criterion measure, "Typology of Creativity." Furthermore, such art task products could bias the selection of students from multiple classes or schools for special art classes if this criterion measure served as a basis for such selection.

Due to negative interjudge correlations between the two classroom teachers who comprised the Teacher Judges, their further value in this study was highly questionable. It was decided that the Teacher Judges would not play any significant part in this study.

CHAPTER IV

RESULTS OF EXPERIMENT

This chapter reports the conclusions reached about each of the primary and secondary hypotheses which form the basis of this study. It would be well, however, to preface the consideration of the results obtained from testing the individual hypotheses by reviewing briefly the nature of the two criterion measures used in this study and the basic questions which underlie the formation of the stated hypotheses.

The Typology of Creativity measure directs a judge to attend to the delineated subject matter and forms of the art product in determining the degree of creativeness of the art work; whereas, the Six Elements of Spontaneity directs a judge to attend to the surface treatment of the art product to determine the degree of spontaneity exhibited in the execution of the art work. Spontaneous treatment of art work surface is considered to identify a working strategy that may lead to creative results through discovery. In addition to the purpose of testing and comparing these two measures within the defined parameters of the present study, there is the ancillary interest in the Six Elements of Spontaneity measure as to its utility in identifying and measuring the spontaneous work strategy in art products of subjects younger than the subjects used in previous studies incorporating this criterion measure. The authors of this criterion measure had implied that the quality of an art production exists independently of the creative spontaneous strategy, and thus,

the quality of the art production should not operate as an intervening variable in the identification and measurement of the employed strategy at any age level.

The hypotheses were formed to test three basic questions. Of the two primary hypotheses, Hypothesis I was formed to test the question of the degree of correlation between the two criterion measures in identifying the creativeness of art products of the same population of subjects. Hypothesis I attends to, first of all, the degree of correlation between the two rank-orders obtained from employing the criterion measures in the evaluation of products of each art task, and, secondly, the summed average of correlation coefficients obtained from across art task judging draws attention to the degree of association between the evaluation results from employing the two criterion measures in rank-ordering the products of eight art tasks. Hypothesis II, the second primary hypothesis, attends to the consistency of creative performance of the subjects as reflected by the ranking of their art products across all possible art task pairings on the basis of each criterion measure. Although there is ample evidence that creative performance is not a consistent kind of behavior in terms of degree, it is worth knowing whether, and to what extent, there is a difference in performance levels as measured by the two criteria because each of the criterion measures attends to different aspects of the art product. For example, the Six Elements of Spontaneity attend to the art product as a reflection of the procedural strategy in the art-making act. It seems plausible that the spontaneous work

strategy employed by a subject could be reflected at a fairly consistent degree among different art tasks; yet, the creativeness of the various art products would fluctuate extensively as measured by the Typology of Creativity measure.

Hypotheses III, IV, and V, which are the secondary hypotheses, were formed to test the question of whether there are particular systematic causes of variance of creative performance in art making. Three variables were inductively devised from the review of studies in art education which seemed to hold promise as sources of systematic variance in creative artistic behavior: (1) space dimensionality; (2) resistance level of art materials; and (3) sensorial empathy with art materials. The variables served as a basis for dichotomizing the products of the art tasks into two defined categories for each variable being tested. First of all, the summed average correlation coefficient was computed from all possible rank-order pairings between the two categories of each variable to determine if there was a statistically significant difference between the rank-orders of products of art tasks of the two categories. Secondly, the summed average correlation coefficient was computed from all possible rank-order pairings within each category to determine if there was a statistically meaningful similarity among the rank-orders of products of art tasks included within each specific category. Any meaningfully significant correlation between the rank-orders of products of art tasks within any one or more categories of the three dichotomies would lend evidence to the hypothesis that there are variables which operate

systematically in determining the consistency of creative artistic behavior as defined by the two criterion measures.

Space dimensionality as a source of systematic variance was tested by Hypothesis III, and the art tasks and products were dichotomized into two- and three-dimensional categories for this purpose. Hypothesis IV was formed to determine if the level of resistance to modification of different art media would act as a source of systematic variance in "creative" and "spontaneous" art performance of the subjects under study. For this purpose, art tasks and products were dichotomized by the degree of inertness or fluidity of the art media when manipulated by the subjects during a particular art task. For instance, the traditional use of watercolors is characterized by the fluid quality of the medium in its working state. On the other hand, clay could be classified as either fluid or inert in its working state; that is, modeling with clay in its moist state would be relatively fluid whereas the carving of clay is done most successfully at the leatherhard stage (inert) which does not lend itself easily to modifications or corrections. To test Hypothesis V, which was formed to determine the degree that sensorial empathy with the art media operated as a source of systematic variance, the art tasks and products were dichotomized into indirect and direct sensorial contact with the art media categories. Compared to the other two dichotomies in which the categorizing of products of art tasks was determined by the relative degree that the products were identified with a particular category, the identification of the products of the art tasks as belonging to one of the categories of

this third dichotomy was straightforward; that is, the subjects either used or did not use tools in forming their art products.

The testing of Hypothesis VI was not completed. The question underlying this hypothesis was whether the teachers with an intimate knowledge of the subjects as art students could, without prior training in the use of the criterion measures, arrive at rank-orders of their students' "creativeness" and "spontaneity" which would be similar to the ranking of subjects as reflected by their art products by the Typology Judges (creativity) and the Six Criterion Judges (spontaneity). However, high negative correlations between the rank-order results by the two teachers, who composed the Teacher Judges group, on both criterion measures would have made the further pursuance of the testing of this hypothesis meaningless.

The hypotheses are stated in terms of predicting significant differences between the rank-ordering of art products on each hypothesis tested. In general, the stated hypotheses reflect the expected results from testing the hypotheses.

For testing the hypotheses, the Spearman Rank Correlation Coefficient: rs was used throughout. This statistic, sometimes called rho, is here represented by 'rs.' It is a measure of association (correlation) which requires that both variables will be measured in at least an ordinal scale so that the art products under study may be ranked in two ordered series. In reporting the results of the hypotheses testing, reference will be made to 'rs' value (correlation coefficient) as a means of indicating the degree of association

between two rank-orders of art products. Reference to 'rs_{av}' value indicates the summed average of 'rs' values among paired rank-orders.

An 'rs' value of .45, or greater, was required to surpass the .10 level of confidence and any value as great as .45 will be noted and discussed for its suggestive implications. However, it is one thing to accept a significant correlation as being real and another to accept confidence in its usefulness. For instance, a correlation coefficient of .45 achieved in this study would be relatively small in spite of its significance. For this reason and due to possible contributing sources of inaccuracies in this study, it was arbitrarily decided that an 'rs' value would have to be as great as .70 before any confidence could be put into its usefulness.

HYPOTHESIS I

THERE WILL BE A STATISTICALLY SIGNIFICANT DIFFERENCE BETWEEN THE POOLED RANKINGS OF THE TYPOLOGY JUDGES AND THE SIX CRITERION JUDGES ON EACH ART TASK AND ACROSS ALL ART TASKS

Results and Observations of the Testing of Hypothesis I

Significant differences were expected between the rank-orders of art products by the two judge groups on each art task and, in particular, a substantial variety of 'rs' values were expected across art tasks. These expectations were borne out in the testing of Hypothesis I. The results are summarized in Table V.

TABLE V

'rs' VALUES OBTAINED BETWEEN THE TYPOLOGY JUDGES AND THE SIX CRITERION JUDGES ON EACH ART TASK AND ACROSS ALL ART TASKS

Art task	(N)	rs
1. Cartoons	22	-.033
2. Prints	21	.469***
3. Ceramic designs	20	.497***
4. Ceramics	16	.306
5. Sculpture	17	.278
6. Papier mâché designs	16	.490****
7. Papier mâché plaques	16	.537**
8. Fantasy drawings	19	.673*
	rs_{av}	.402

*Significant at the .01 level of confidence
**Significant at the .02 level of confidence
***Significant at the .05 level of confidence
****Significant at the .10 level of confidence

Although five of the eight 'rs' values were significant at the .10 level of confidence or better, no 'rs' value reached the prescribed .70 degree of association required to be considered meaningful. Furthermore, and not surprisingly, the 'rs_{av}' value of .402 across art tasks is neither significant or meaningful. A possible contributing factor to the obtained variance between the evaluation results of employing the two criterion measures is that the "Six Elements of

Spontaneity" measure identifies only one of possibly many creative work strategies. For instance, there is the working strategy defined by Beittel and Burkhart as "divergent," i.e., ". . . a creative strategy which proceeds through deliberation toward discovery."[1]

The variance of obtained 'rs' values across art tasks suggests the possibility that there may be some art tasks that bring out and intensify those qualities which are identifying characteristics of both criterion measures, and these qualities are equal, but different, in their ability to identify artistic creativity. Although the results of treating Hypothesis I do not prove or disprove such a possibility, none of the art tasks as presented to the subjects in this study resulted in a meaningful correlation between the rank-orders of their art products by the two judge groups. The correlation between the rank-orders of the products of the fantasy drawings task obtained the highest 'rs' value (.673) of the eight art tasks and is worthy of further consideration.

The 'rs' value of .67 obtained between the two judge groups' rank-order evaluation of the products of the fantasy drawings task is significant at the .01 level of confidence and nearly satisfies the specified 'rs' value of .70 required for a meaningful interpretation and conclusion. Even if the 'rs' value had surpassed the specified 'rs' value of .70, which it may have done in another judging situation,

---

[1]Kenneth R. Beittel and R. C. Burkhart, "Strategies of Spontaneous, Divergent, and Academic Art Students," Studies in Art Education, 5:20-41, No. 1, (1963), p. 20.

the employment of the fantasy drawing task as a valid stimulus for soliciting from subjects their optimal potential for artistic creativity as defined by both criterion measures is questionable. It is questionable whether there is a meaningfully significant correlation between the manifestations of fantasy and any of the multiple definitions of creativity or imagination. The fantasy drawing task may have value in achieving particular objectives of art learning behavior; however, it is doubtful that the products of the fantasy art task would reflect the insight and imaginative interpretation of a subject's personal, social, and physical environment which is so much a part of recognized creative, artistic efforts. What the products of the fantasy art task might do is reflect a subject's psychological distance from reality.

The reality which needs to be reemphasized is the results of testing Hypothesis I, which show that there were no 'rs' values obtained between the rank-orders of the two judge groups on any art task or across art tasks which could be considered meaningful. To consider the possibility that "spontaneous" or "creative" behavior is consistently displayed by the comparison of rank-orders of products among art tasks, we turn to the testing of Hypothesis II.

HYPOTHESIS II

THERE WILL BE A STATISTICALLY SIGNIFICANT DIFFERENCE AMONG THE POOLED RANKINGS ASSIGNED TO THE PRODUCTS OF DIFFERENT ART TASKS BY THE TYPOLOGY JUDGES AND THE SIX CRITERION JUDGES

In comparison to Hypothesis I, in which the problem was to

determine whether both judge groups would arrive at a similar rank-order on each art task and across all art tasks, the problem in Hypothesis II was to determine whether each of the two judge groups would arrive at similar rank-orders among art tasks.

In the testing of Hypothesis II and succeeding hypotheses, it was necessary in many instances to delete some subjects' art product rank from consideration. This would, in some instances, change the numerical rank of the subjects' art products but would not change their position in a particular art product rank-order.

Results and Observations of the Testing of Hypothesis II

Tables VI and VII are matrices of the intercorrelation coefficients obtained among paired art-task ranks from each judge group. Of particular significance in the testing of Hypothesis II are the 'rs_{av}' values obtained by the Six Criterion Judge group and the Typology Judge group (.031 and .035 respectively). These near-zero correlations indicate, statistically, the high degree of variance among rank-orders assigned to the subjects' products of the different art tasks.

Accepting that all other variables had remained constant during the time-period of this study, then the source of variance is probably in the nature of the art task and materials. The fact is that variance, and not consistency, characterizes the performance of the subjects as reflected by their art product rankings. The art products, as a reflection of the subjects art performance, indicate that the art task and media played no small part in determining the

TABLE VI

MATRIX OF INTERCORRELATIONS AMONG PAIRED ART TASK RANKS
FROM THE EVALUATION RESULTS OF THE SIX CRITERION JUDGES

		1	2	3	4	5	6	7	8
1.	Cartoons		.143	-.410***	-.244	.144	-.074	-.313	.037
2.	Prints			.033	-.180	.274	-.119	-.324	-.266
3.	Ceramic designs				.536**	-.131	.081	.324	.087
4.	Ceramics					-.205	.557**	.557*	.171
5.	Sculpture						-.335	-.187	.318
6.	Papier mâché designs							.383	-.245
7.	Papier mâché plaques								.284
8.	Fantasy drawings			('rs_{av}' .031)					

*Significant at the .05 level of confidence
**Significant at the .10 level of confidence
***Negatively significant at the .10 level of confidence

TABLE VII

MATRIX OF INTERCORRELATIONS AMONG PAIRED ART TASK RANKS
FROM THE EVALUATION RESULTS OF THE TYPOLOGY JUDGES

		1	2	3	4	5	6	7	8
1.	Cartoons		.040	.044	.231	-.316	.357	.018	.151
2.	Prints			-.157	-.181	.208	.171	-.351	.029
3.	Ceramic designs				.071	.175	-.168	.366	-.236
4.	Ceramics					.278	.127	-.191	-.210
5.	Sculpture						.416	.141	.081
6.	Papier mâché designs							.012	.044
7.	Papier mâché plaques								-.171
8.	Fantasy drawings			('rs_{av}' .035)					

degree of "creativeness" of an individual's art work and how "spontaneously" he would handle the art materials in executing the art product.

To say, however, that no subject displayed consistent "creative" or "spontaneous" behavior would be incorrect. It is not known whether some or all subjects varied in their performance across art tasks; or put another way, an individual's change of rank does not necessarily indicate a change in performance.

The contention of some researchers that a single art task, form, and medium can be used to determine children's creative or artistic ability in the visual arts must be seriously questioned. For example, the use of crayon, or any medium for that matter, as being the most appropriate medium for executing the product of any art task test has its limitations. The specification of a particular art medium because it is the most familiar medium to the subjects disregards the quality of experiences that the subjects have had with the medium and their personal reactions toward it. Both criterion measures used in this study are derived from and/or applied to limited, specified art tasks and materials. The "Six Criterion Elements of Spontaneity" was developed from the analysis of drawings of complex still-life problems executed in pen-and-ink and brush. That is, both subject matter (stimulus) and medium were held constant as a basis for control and comparison. The other criterion measure, "Typology of Creativity," was composed of two art tasks: (1) a series of drawings in pencil employing a few abstract lines as stimuli, and (2) a

construction employing one-quarter pound of oil base clay, colored toothpicks, and a paper plate for a base which would allow the art work to be easily moved. In the development of this criterion measure, art materials were controlled but not the subject matter.

Whether the subject matter is specified or the art materials are specified, the question remains, do all subjects, or even the majority of subjects, in the test situation identify with the tasks to the extent that their creative ability becomes manifest in the test products?

In sum, any of the art tasks and media included in this study or those art tasks and media which were incorporated in the development of the two criterion measures could be questioned as to their validity as appropriate vehicles for determining creative ability. Further evidence will be presented in this chapter which will cast doubt on those studies of researchers who assume that there are certain art tasks and media which are most appropriate for visually soliciting one's creative potential.

The range of variation of rank-orders among the products of the art tasks requires pursuance of the systematic source(s) of this variance, and this entails a comparison of art task results which, in part, is dealt with in the testing of the secondary hypotheses.

SECONDARY HYPOTHESES

The secondary hypotheses include Hypotheses III, IV, and V and were inductively devised to test for possible sources of variance.

Primarily, they were devised on the basis of variables whose dichotomies would separate art tasks as to space dimensionality, resistance of art materials, and sensorial empathy with art materials. The possibility that variables would be discovered which operate as systematic sources of variance was foredoomed to failure by the results obtained from testing Hypothesis II. It will be noted in the perusal of the 'rs' values reported in Tables VI and VII that none surpassed the specified .70 degree of correlation for meaningful interpretation or conclusion. However, the testing of the secondary hypotheses gives insight into some of the difficulties of identifying systematic variables which operate in art production.

In the testing of the secondary hypotheses two kinds of statistical computations were made. First of all, art tasks were dichotomized into the two defined categories of the variable being tested and then all possible rank-order pairings between the two categories were correlated. The averaging of the obtained correlation coefficients as reflected by the resulting 'rs_{av}' value would determine if there was a statistically significant difference between the rank-orders of products of art tasks of the two categories.

Secondly, all possible pairings of rank-orders within each category were correlated to determine if there was a statistically meaningful similarity among the rank-orders of products of art tasks which comprised a category. Again, the averaging of the obtained correlation coefficients as reflected by the resulting 'rs_{av}' value would indicate the degree of similarity among these rank-orders.

Any 'rs_{av}' which surpassed the .70 degree of correlation would mean that the defining characteristic of that category operates as a systematic source of variance in creative artistic behavior as defined by the criterion measure used for evaluation.

Statistically significant differences were expected between the rank-orders of art task products which comprised the two categories of each dichotomy. However, high positive correlations, if indeed not statistically significant and meaningful correlations, were expected among the rank-orders of art tasks which comprised any one category.

Art Tasks

The art tasks will be at times referred to by number, and the following list of art tasks not only gives the identification number of each art task, but also is arranged in the sequence of the art task completion.

Number	Art task
1.	Cartoons
2.	Prints
3.	Ceramic designs
4.	Ceramics
5.	Sculpture
6.	Papier mâché plaque designs
7.	Papier mâché plaques
8.	Fantasy drawings

HYPOTHESIS III

THERE WILL BE A STATISTICALLY SIGNIFICANT DIFFERENCE BETWEEN THE POOLED RANKINGS OF THE SUBJECTS' TWO- AND THREE-DIMENSIONAL ART PRODUCTS. THIS DIFFERENCE WILL BE FOUND BETWEEN THE POOLED RANKINGS OF EACH JUDGE GROUP.

There are variables which are inherently related to each two- and three-dimensional art performance, and the question pertaining to this hypothesis is whether these variables operate to such an extent as to be a source of systematic variance across art tasks.

Results and Observations of the Testing of Hypothesis III

The results summarized in Table VIII clearly reflect the difference between the rank-orders of products of two- and three-dimensional art tasks. The 'rs_{av}' values obtained for each judge group are very low and, in fact, approach a zero relationship.

The 'rs' values obtained among paired art task rankings range from significantly positive correlations to nearly significantly negative correlations. There are only two instances among paired rankings in which significant correlations between two- and three-dimensional art product rank-orders were obtained. Both of these correlations were obtained from the evaluation results of the Six Criterion Judges, and both of them were related to the measure of spontaneity of ceramic products (art task 4) in their relationship to the spontaneous rankings of 1) ceramic designs (art task 5) and 2) papier mâché plaque designs (art task 6). To explain the obtained correlation between the rankings of cermaic design products and ceramic products as the latter being the outgrowth of the first is simple enough; however, it would make more sense if such a positive correlation also had existed in the evaluation results of the Typology Judges. There is no more reason for the obtained correlation between

the ranks of ceramic products and papier mâché plaque design products than for a significantly positive correlation between the ceramic task and any other task. There is no triangulation of positive significant correlations between these three art tasks since the 'rs' value obtained between ceramic design products and papier mâché design products' ranking was .081 on the spontaneity measure, and the best that can be said of this 'rs' value is that it is positive.

TABLE VIII

'rs' VALUES OBTAINED BETWEEN PAIRINGS OF TWO- AND THREE-DIMENSIONAL ART TASK PRODUCT RANKINGS

	Art task pairings	rs(N)	Art task pairings	rs(N)	Art task pairings	rs(N)
Typology Judges	4-1	.231(16)	5-1	-.316(17)	7-1	.018(19)
	4-2	-.181(15)	5-2	.208(16)	7-2	-.351(18)
	4-3	.071(14)	5-3	.175(15)	7-3	.366(17)
	4-6	.127(11)	5-6	.416(15)	7-6	.012(15)
	4-8	-.210(12)	5-8	.018(14)	7-8	-.171(15)
					('rs_{av}'	.032)
Six Criterion Judges	4-1	-.244(16)	5-1	.144(17)	7-1	-.313(19)
	4-2	-.180(15)	5-2	.274(16)	7-2	-.324(18)
	4-3	.536(14)*	5-3	-.131(15)	7-3	.324(17)
	4-6	.557(11)**	5-6	-.335(15)	7-6	.383(15)
	4-8	.171(12)	5-8	.318(14)	7-8	.284(15)
					('rs_{av}'	.098)

*Significant at the .05 level of confidence
**Significant at the .10 level of confidence

It is clearly accepted that differences between the rank-orders of two- and three-dimensional art products is statistically significant; however, to say that the variables which are inherent to each of the

two- and three-dimensional art tasks operate as a systematic source of variance across art tasks must be rejected. By computing the 'rs' values between the rank-order pairings of the three art tasks whose products were three-dimensional, by definition, a large degree of variation is noted.

Table IX shows that the obtained 'rs_{av}' values between paired rankings of three-dimensional art task products is very low for each judge group, and the range of obtained 'rs' values between paired rankings ranges from positive to negative correlations.

TABLE IX

'rs' VALUES OBTAINED BETWEEN RANKINGS OF THREE-DIMENSIONAL ART PRODUCTS

	Art task pairings	rs(N)	Art task pairings	rs(N)	Art task pairings	rs(N)
Typology Judges	4-5	.278(12)	4-7	-.191(14) ('rs_{av}' .076)	5-7	.141(17)
Six Criterion Judges	4-5	-.205(12)	4-7	.557(14)* ('rs_{av}' .055)	5-7	-.187(17)

*Significant at the .05 level of confidence

The only significantly positive correlation was obtained between the rank-orders of the art products from the ceramic (art task 4) and papier mâché plaques (art task 7) on the spontaneity measure. This is a reasonable correlation because both papier mâché and clay materials are bulky and lend themselves more to large, freely executed products than other art materials employed in the other art tasks. Coupled

with the lack of dexterity on the part of the subjects in manipulating these art materials, the products of both art tasks were characterized by a spontaneous-like quality.

The sculpture products (art task 5), although done with clay, were completed, in a large part, by using tools; whereas, the ceramic products were hand modeled. The sculpture products were characterized more by a "hard-edge" rather than a spontaneous-like quality.

Statistically, it must be concluded that three-dimensionality is not a systematic variable among art task rankings.

The obtained 'rs' values between paired ranks of the art tasks whose products were two-dimensional, by definition, are also notable for the degree of variation among them. The 'rs_{av}' values were again very low and not significant. Table X shows that no significantly positive correlation was obtained.

TABLE X

'rs' VALUES OBTAINED BETWEEN RANKINGS OF TWO-DIMENSIONAL ART PRODUCTS

	Art task pairings	rs(N)	Art task pairings	rs(N)
Typology Judges	1-2	.040(21)	2-6	.171(15)
	1-3	.044(20)	2-8	.029(15)
	1-6	.357(16)	3-6	-.168(14)
	1-8	.151(16)	3-8	-.236(14)
	2-3	-.157(19)		
			('rs_{av}' .077)	
Six Criterion Judges	1-2	.143(21)	2-6	-.119(15)
	1-3	-.410(20)	2-8	-.266(15)
	1-6	-.074(16)	3-6	.081(14)
	1-8	.037(16)	3-8	.087(14)
	2-3	.033(19)		
			('rs_{av}' -.163)	

Variables which exist between two- and three-dimensional art tasks may contribute to rank variation across art tasks, but it is likely that a major portion of the variance is due to other factors.

HYPOTHESIS IV

THERE WILL BE A STATISTICALLY SIGNIFICANT DIFFERENCE BETWEEN THE POOLED RANKINGS OF THE SUBJECTS' ART PRODUCTS EXECUTED WITH FLUID ART MEDIA AND THOSE EXECUTED WITH INERT ART MEDIA. THIS DIFFERENCE WILL BE FOUND BETWEEN THE POOLED RANKINGS OF EACH JUDGE GROUP.

The thinking here is that a fluid medium would be compatible to a "spontaneous" type student, and an inert art medium would be more attractive to the meticulous subject. If this were so, then the fluid or inert quality of the art media used would be a source of systematic variance across art task products' ranking.

Results and Observations of the Testing of Hypothesis IV

The results of testing Hypothesis IV are summarized in Table XI. Correlations obtained between paired ranks of art task products executed in fluid and inert art media are extremely low with an 'rs_{av}' of .046 obtained from the evaluation results of the Typology Judges and a 'rs_{av}' value of .123 obtained from the evaluation results of the Six Criterion Judges. As in the testing of Hypothesis III, the variety of 'rs' values extends from nearly significant negative correlations to significantly positive correlations of which none surpassed the .70 degree of correlation for meaningful interpretations or conclusions.

TABLE XI

'rs' VALUES OBTAINED BETWEEN PAIRED RANKS OF ART TASKS EXECUTED IN FLUID AND INERT ART MEDIA

	Art task pairings	rs(N)	Art task pairings	rs(N)
Typology Judges	4-1	.231(16)	5-1	-.316(17)
	4-2	-.181(15)	5-2	.208(16)
	4-3	.071(14)	5-3	.175(15)
	4-6	.127(11)	5-6	.416(15)
	4-7	-.191(14)	5-7	.141(17)
	4-8	-.210(12)	5-8	.081(14)
			('rs_{av}' .046)	
Six Criterion Judges	4-1	-.244(16)	5-1	.144(17)
	4-2	-.180(15)	5-2	.274(16)
	4-3	.536(14)*	5-3	-.131(15)
	4-6	.557(11)**	5-6	-.335(15)
	4-7	.557(14)*	5-7	-.187(17)
	4-8	.171(12)	5-8	.318(14)
			('rs_{av}' .123)	

*Significant at the .05 level of confidence
**Significant at the .10 level of confidence

The three 'rs' values which were statistically significant are again found within the evaluation results of the Six Criterion Judges. Art tasks 3, 4, 6, and 7 do have some common characteristics which have suggestive implications; however, they will be considered later in this chapter because they are not directly relevant to the testing of this hypothesis.

Again, as in Hypothesis III, variables which exist between the executing of art works in inert and fluid art media may be a source of rank variance across art tasks, but it is quite likely that a major portion of the variance is due to other factors. Table XII shows

that no significant positive correlation was obtained from the paired ranks of the art task products done in a fluid art medium; that is, the fluidity of a medium alone is not a systematic cause of the observed variance. However, it needs to be pointed out that the medium used in both the ceramic and sculpture art tasks was clay, and other fluid art media may produce different results.

TABLE XII

'rs' VALUES OBTAINED BETWEEN PAIRED RANKS OF ART TASKS WHOSE PRODUCTS WERE DONE IN A FLUID ART MEDIUM (ART TASKS 4 AND 5)

Typology Judges	Six Criterion Judges
.278	-.205

Table XIII shows that the obtained 'rs' values between paired ranks of art tasks whose products were executed with inert art media, by definition, are also notable for the degree of variation among them. The 'rs_{av}' values obtained from each judge group clearly indicate that there is a near zero relationship among paired ranks of art tasks whose products were executed with inert art media. There were no significant positive correlations, but there was one significant negative correlation. This correlation was obtained between the paired ranks of the cartoon and ceramic design art task' products and within the evaluation results of the Six Criterion Judges. Table XIII shows that the rank-order of products of this art task obtained either

low or negative correlations with all other art task product ranks. This holds true for the two art tasks for which no correlations from paired ranks with the cartoon task were listed in Table XIII (Art tasks 1 and 4 obtained the 'rs' value of -.244 and the obtained 'rs' value between art task 1 and 5 was .144). It has also been noted in the testing of Hypothesis I, in which the degree of relationship between the Typology Judges and the Six Criterion Judges on each art task and across art tasks was determined, that the cartoon task was the only instance in which a negative correlation was obtained.

TABLE XIII

'rs' VALUES OBTAINED BETWEEN PAIRED RANKS OF ART TASKS WHOSE PRODUCTS WERE EXECUTED WITH INERT ART MEDIA

	Art task pairings	rs(N)	Art task pairings	rs(N)
Typology Judges	1-2	.040(21)	3-6	-.168(14)
	1-3	.044(20)	3-7	.366(17)
	1-6	.357(16)	3-8	-.236(14)
	1-7	.018(19)	6-7	.012(15)
	1-8	.151(16)	6-8	.044(13)
	2-3	-.157(19)	7-8	-.171(15)
	2-6	.171(15)		
	2-7	-.351(18)	('rs_{av}' .010)	
	2-8	.029(15)		
Six Criterion Judges	1-2	.143(21)	3-6	.081(14)
	1-3	-.410(20)*	3-7	.324(17)
	1-6	-.074(16)	3-8	.087(14)
	1-7	-.313(19)	6-7	.383(15)
	1-8	.037(16)	6-8	-.245(13)
	2-3	.033(19)	7-8	.284(15)
	2-6	-.119(15)		
	2-7	-.324(18)	('rs_{av}' -.027)	
	2-8	-.266(15)		

*Negatively significant at the .10 level of confidence

The reason for such a high negative correlation seems to be attributable to the nature of the cartoon art task. Of all art tasks, it would probably be safe to say that the products of this art task were the most stereotyped; the subjects depended upon popular cartoon drawings; and in drawing their cartoons, there seemed to be much effort on their part to render the cartoons as faithfully to the "original" as possible. The results were generally low on "spontaneity." Although this may be a reasonable explanation, it was not statistically proved.

In summary, the variables which exist between art products executed with fluid art media and those with inert art media may be a source of rank variance across art tasks, but it is quite likely that a major portion of the variance is due to other factors.

HYPOTHESIS V

THERE WILL BE A STATISTICALLY SIGNIFICANT DIFFERENCE BETWEEN THE POOLED RANKINGS OF THE SUBJECTS' ART PRODUCTS WHICH ENTAILED DIRECT SENSORIAL CONTACT WITH THE ART MEDIA AND THOSE WHICH ENTAILED INDIRECT SENSORIAL CONTACT WITH THE ART MEDIA. THIS DIFFERENCE WILL BE FOUND BETWEEN THE POOLED RANKINGS OF EACH JUDGE GROUP.

The rationale underlying this hypothesis is that a greater degree of empathy exists with a medium which is handled directly; that is, a subject is more likely to go with the medium rather than to make it conform to any preconceived strategy or purposes of employing that particular medium. This characteristic of being "in-tune" with the medium would seem to be more compatible with a

spontaneous-type person and would operate as a source of variance across art tasks and in the evaluation results from the employment of both criterion measures.

There were two art tasks in which the subjects used their hands to execute the final product: the products of the ceramic and papier mâché art tasks. Sculpture was not included because the subjects employed ceramic tools quite extensively to complete their products.

Results and Observations of the Testing of Hypothesis V

Table XIV shows that there is quite a range of variation between rankings of art task products which required direct sensorial contact with the medium and those in which tools were used in executing the art products.

The obtained 'rs_{av}' values from the evaluation results of each judge group again approaches zero. The only two 'rs' values which resulted in significantly positive correlations are found within the computed correlations from the evaluations of the Six Criterion Judges. Art tasks 3, 4, and 6 are three of four art tasks which were noted in the testing of Hypothesis IV as having possible suggestive implications (the other art task was number 7). These tasks will be considered later in relation to the significantly positive correlation between the two art tasks which required direct sensorial contact with the art media.

There is clearly a difference between the rank-orders of those art products in which there was direct sensorial contact with the

TABLE XIV

'rs' VALUES OBTAINED BETWEEN PAIRED RANKS OF ART TASKS WHICH REQUIRED DIRECT SENSORIAL CONTACT WITH THE MEDIUM AND THOSE ART TASKS IN WHICH TOOLS WERE EMPLOYED

	Art task pairings	rs(N)	Art task pairings	rs(N)
Typology Judges	4-1	.231(16)	7-1	.018(19)
	4-2	-.181(15)	7-2	-.351(19)
	4-3	.071(14)	7-3	.366(17)
	4-5	.278(12)	7-5	.141(17)
	4-6	.127(11)	7-6	.012(15)
	4-8	-.210(12)	7-8	-.171(15)
			('rs_{av}' .028)	
Six Criterion Judges	4-1	-.244(16)	7-1	-.313(19)
	4-2	-.180(15)	7-2	-.324(18)
	4-3	.536(14)*	7-3	.324(17)
	4-5	-.205(12)	7-5	-.187(17)
	4-6	.557(11)**	7-6	.383(15)
	4-8	.171(12)	7-8	.284(15)
			('rs_{av}' .067)	

*Significant at the .05 level of confidence
**Significant at the .10 level of confidence

medium and those executed with tools, but as with the results obtained from testing Hypotheses III and IV, the possibility that the hypothesized variables operate as a source of systematic variance across art tasks is questionable. The results from computing 'rs' values between paired ranks of art tasks in which tools were employed in manipulating art media are summarized in Table XV. The obtained 'rs_{av}' values for each judge group clearly indicate that there is a near zero relationship between paired ranks among these art tasks.

TABLE XV

'rs' VALUES OBTAINED BETWEEN PAIRED RANKS OF ART TASKS IN WHICH TOOLS WERE EMPLOYED IN MANIPULATING ART MEDIA

	Art task pairings	rs(N)	Art task pairings	rs(N)
Typology Judges	1-2	.040(21)	2-8	.029(15)
	1-3	.044(20)	3-5	.175(15)
	1-5	-.316(17)	3-6	-.168(14)
	1-6	.357(16)	3-8	-.236(14)
	1-8	.151(16)	5-6	.416(15)
	2-3	-.157(19)	5-8	.081(14)
	2-5	.208(16)	6-8	.044(13)
	2-6	.171(15)		
			('rs_{av}' .056)	
	1-2	.143(21)	2-8	-.266(15)
	1-3	-.410(20)*	3-5	-.131(15)
	1-5	.144(17)	3-6	.081(14)
	1-6	-.074(16)	3-8	.087(14)
	1-8	.037(16)	5-6	-.335(15)
	2-3	.033(19)	5-8	.318(14)
	2-5	.274(16)	6-8	-.245(13)
	2-6	-.119(15)		
			('rs_{av}' .055)	

*Negatively significant at the .10 level of confidence

Nearly all correlations in Table XV are repetitions from earlier reports. The negative correlation obtained between the rank-order of the cartoon art task (1) and the ceramic designs art task (3) was considered previously to be attributable to the characteristics of the cartoon art task' products.

Table XVI shows that by computing the correlations between the rank-orders of art tasks which required direct sensory contact with art media, a low, negative correlation was obtained from the evaluation

results of the Typology Judges; whereas, a significantly positive correlation was obtained from the evaluation results of the Six Criterion Judges.

TABLE XVI

'rs' VALUES OBTAINED BETWEEN PAIRED RANKS OF ART PRODUCTS WHICH REQUIRED DIRECT SENSORY CONTACT WITH THE ART MEDIA (ART TASKS 4 AND 7)

Typology Judges	Six Criterion Judges
-.191	.557*

*Significant at the .05 level of confidence

In spite of the significantly positive correlation obtained between the rank-orders of the ceramics art task and the papier mâché art task from the evaluation results of the Six Criterion Judges, it is doubtful that direct sensorial contact with the art media is any more a source of systematic variance across art tasks than the hypothesized systematic variables tested in Hypotheses III and IV. It may well be that, if more than two samples of art products which required direct sensory contact with the art media had been available for comparison, a significantly positive correlation among such art task rank-orders would not have been obtained. To justify this contention, the obtained significantly positive correlation between art product ranks of the ceramics and papier mâché art tasks must be elaborated upon.

The correlation between art tasks 4 and 7 and the obtained correlations between these two art tasks and art task 3 (ceramic designs) and 6 (papier mâché plaque designs) were noted in the testing of Hypotheses III and IV. It was pointed out at that time that the evaluation results of the products of these four art tasks by the Six Criterion Judges offered some suggestive implications. Art tasks 3, 4, 6, and 7 have some related characteristics: (1) art tasks 3 and 6 were designs for art tasks 4 and 7, (2) art tasks 4 and 7 involved three-dimensional problem-solving, (3) art tasks 3 and 6 were done with crayon, and (4) art tasks 4 and 7 required direct sensorial contact with the art media.

Table XVII summarizes the results of selecting these four art tasks and extending the pairings among them along with the obtained 'rs' values. The evaluation results of the Typology Judges are also included for comparison. The extension of the significance level to .20 is introduced only to indicate that particular 'rs' values were large enough to achieve this level of confidence. It is interesting that the extension of rank-order pairings in Table XVII includes all 'rs' values which surpassed the .20 level of confidence.

Accepting for the present that valid conclusions can be drawn from these results, four characteristics of these art tasks can be elaborated upon. Looking at the 'rs' values obtained by the Six Criterion Judges, first of all, Table XVII shows that the two designs done in crayon (art tasks 3 and 6) have the lowest correlation (.081). These tasks have two common characteristics: medium and design problem.

TABLE XVII

'rs' VALUES OBTAINED BETWEEN PAIRED RANKINGS OF ART TASKS 3, 4, 6, AND 7

	Art task pairings	rs(N)	Art task pairings	rs(N)
Typology Judges	3-4	.071(14)	6-7	.012(15)
	3-6	-.168(14)	4-6	.127(11)
	3-7	.366(17)***	4-7	-.191(14)
Six Criterion Judges	3-4	.536(14)*	6-7	.383(15)***
	3-6	.081(14)	4-6	.557(11)**
	3-7	.324(17)	4-7	.557(14)*

*Significant at the .05 level of confidence
**Significant at the .10 level of confidence
***Significant at the .20 level of confidence

The conclusion which could be drawn is that familiarity with crayon or design alone is no guarantee of consistency of spontaneity levels in the subjects' art products. Crayon had been used previously, not only in conjunction with art task 1, but also before the inception of the study. If familiarity with the medium were a major determiner of consistency of surface treatment then no matter how stereotyped or how spontaneous subjects were in using the medium, their rankings across art tasks employing the medium crayon would show little variation. Familiarity with designing must be approached more cautiously. For one thing, there was little previous experience with design, and this was, in the main, limited to elementary school experiences. Secondly, the term "design" is used loosely as the nature of the problems presented to the students was more "planning" than designing. What

instruction was given in design was almost entirely on an individual basis and on immediate need. It is questionable that the subjects had any real understanding of design and/or depth in the experience of designing. For this reason, the understanding and skill of designing as an influencing variable on the consistency of spontaneity levels of performance cannot lead to the suggestive implication which was stated in relation to the crayon art medium.

Continuing from the position that valid conclusions can be drawn from the information summarized in Table XVII, the 'rs' values obtained by the Six Criterion Judges between the pairings of "before/after" art tasks (3-4 and 6-7) were fairly high; however, the 'rs' value obtained between papier mâché design and the final papier mâché plaque was low reaching the significance level of .20 on the two-tailed test. The most common and dominant characteristic of the "before/after" tasks is the depth of involvement with the subject of their art products which the students had experienced. It could be concluded that it is only at that time when children have confidence and satisfaction in delineating their chosen subject that they will be free in exhibiting their level of spontaneity with consistency. The lower 'rs' value obtained from the art tasks 6-7 pairing could have been due to the lack of the subjects' understanding of the inherent problems which exist in transferring a design done in one medium to a different medium. Crayon, although not entirely compatible to intricately delineated designs, lends itself more readily to detail than the bulky quality of papier mâché. Some of the

subjects in this study discarded their original ideas executed in crayon when they encountered difficulty in transferring their ideas to the papier mâché medium. Although a change of the original idea could have produced a more interesting, more spontaneous art product, they generally produced a product which, due to their frustrations, was dictated more by expediency rather than genuine interest.

The last correlation of interest from the evaluation results of the Six Criterion Judges is the 'rs' value obtained between the two final products of the "before/after" experiences (art tasks 4 and 7). The 'rs' value of .557, although not surpassing the prescribed .70 degree of correlation needed for a meaningful conclusion, was the highest significant correlation obtained between any art task pairing (significant at the .05 level of confidence). The ceramic and papier mâché plaque art tasks required three-dimensional problem-solving. The products were from the only two art tasks which required direct sensorial contact with the art media and the only two art tasks which entailed fairly extensive preplanning by the subjects.

It was concluded from testing Hypothesis III that three-dimensionality does not operate as a systematic variable across art tasks. Although direct sensorial contact with art media may be a source of systematic variance across art tasks, there is a question as to whether more than two samples of art products which required direct sensorial contact with the art media would have produced such a high positive correlation. The possibility that direct sensorial contact with the art media as an influencing variable on the

consistency of spontaneous levels of performance is suggestive enough to warrant further testing under different conditions. Depth of involvement with the theme or presented problem comes on quite strong as a contributing factor which influences the subjects' consistency of spontaneous performance across art tasks. It seems reasonable, however, that depth of involvement must be something more than the length of time spent with a particular art experience before it can be ascertained whether depth experiences actually contribute to consistency of spontaneous performance. On the other hand, it may be that two or more variables such as direct sensorial contact with the art media, three-dimensionality, and depth of exploration of the art problem would contribute to consistent levels of spontaneous performance in art making if the quality of the art experience is meaningful and well-funded informatively.

This may be a reasonable and logical explanation why these particular art tasks pairings resulted in fairly high positive correlations as evaluated by the Six Criterion Judges, but the same explanation is hardly valid in considering the 'rs' values obtained by the Typology Judges. Throughout the testing of Hypotheses III, IV, and V, it was shown that no art task pairing achieved the .10 level of confidence and only one 'rs' value reached the .20 level of confidence on a two-tailed test.

Some of the lower correlations from the evaluation results of the Typology Judges could be due to lower interjudge correlations, but it is also plausible that the difference between the defining characterisitics of the two criterion measures could be a major

factor. It may be that one's level of spontaneity in handling the art materials would be influenced less by unfamiliarity with the medium, subject, or kind of problem than would one's level of creativeness with form or subject.

For one thing, unfamiliarity with the inherent characteristics of a medium can obliterate creative ideas. It has been discussed in consideration of variance in the consistency of spontaneous handling of materials that the transferring of an idea done in one medium to another forces some subjects to modify their original ideas. This observation could hold true of the subjects' inventiveness in subject and form also. Too often novel, inventive concepts of form and subject are covered over because of the lack of skill in handling the medium or there is a lack of compatibility between the medium and the delineated idea. In this sense, the young artist never realizes the potentialities of the medium, only its limitations.

In other instances, the subjects seemed to be concerned with directing their energies toward getting-to-know the subject and not in the novel elaboration of the subject or in exploiting its aesthetic possibilities. For example, this was the case in the sculpture art task. The cartoon art task points out another source of variance between art tasks on the "Typology of Creativity" measure. In this instance, the subjects' concepts of a cartoon were quite stereotyped and, in comparison with other art task products, displayed little in the way of novelty or inventiveness.

These varied orientations by the subjects toward the various art tasks could be a major source of variance in the obtained

correlations among art task products evaluated on the basis of both criterion measures, but most particularly, the "Typology of Creativity" measure. It is well to reiterate that whereas the judging of the "spontaneous" level of an art product is mainly based on surface treatment, the judging of the level of "creativity" is mainly oriented toward the basic structure of an art work, and these underlying characteristics can be hidden by lack of skill in executing the art work in an unfamiliar medium.

No systematic source of variance was identified in the testing of the secondary hypotheses. It is apparent from the range of variance among product ranks of the art tasks that the variables under study in each of the hypotheses, or any other variable, would not operate systematically in the art productions of subjects of a population sample similar to the sample used in this study; however, it is not accepted that the hypothesized variables tested in this study would not operate systematically in the art productions of a different population sample or under different conditions. To be able to say with a high degree of certainty that particular variables do or do not operate systematically in art making, the subjects under study must be assured of having the background of experiences which would optimize consistency of art making performance.

HYPOTHESIS VI

THERE WILL BE A STATISTICALLY SIGNIFICANT DIFFERENCE BETWEEN THE TEACHER JUDGES' POOLED RANKING OF THE SUBJECTS AND THE RANKING OF THE SUBJECTS BY THE TYPOLOGY JUDGES AND THE SIX CRITERION JUDGES. THE

RANKING OF SUBJECTS BY THE TYPOLOGY JUDGES AND THE SIX CRITERION JUDGES WILL BE DETERMINED BY THE POOLED RANKINGS OF THE SUBJECTS' ART PRODUCTS

The question underlying this hypothesis was whether the teachers with an intimate knowledge of the subjects as art students and knowledgeable of their art problem-solving strategies could, without prior training in the use of the criterion measures, arrive at rank-orders of their students which would be similar to the rankings of subjects, as reflected by their art products, by the Typology Judges and the Six Criterion Judges. If any real differences did exist, reasons for these differences were to be sought from the two teachers. Of special interest would be those variables which would be related to the art process rather than the art product.

The teachers were not asked for their evaluations until the collection of all art tasks' products had been completed. They were given abbreviated copies of the defining characteristics of the criterion measures to study. No explanation was offered, but questions were answered. The few questions posed were mainly concerned with the ranking procedures and whether the criterion should be limited to the subjects' art products. The answer to this latter question was that their evaluations should be based on the students and not just the quality of their art products.

The intent of this portion of the study was discarded when negative interjudge correlations were obtained between the two teachers. A significant negative correlation (-.449) at the .05 level

of confidence was obtained between the two teachers' ranking of subjects on the basis of the spontaneity measure, and a significant negative correlation (-.580) at the .01 level of confidence was obtained on the basis of the "Typology of Creativity" measure. The pursuance of the testing of Hypothesis VI would have been wasted effort as the results were predictable and would be meaningless.

SUMMARY OF CONCLUSIONS FROM THE EXPERIMENT

It was shown statistically from the testing of Hypothesis I that the two judge groups employing the criterion measures, "Six Elements of Spontaneity" and the "Typology of Creativity," did not similarly rank-order on the average the products of the eight art tasks ('rs_{av}' .402). The obtained 'rs' values between judge groups on each art task ranged from .673 on fantasy drawings to -.033 on the cartoon art task. None of the correlation coefficients were high enough to surpass the prescribed .70 degree of correlation which was considered necessary before useful conclusions could be drawn. It was suggested that the reason for this range of variation was attributable to the nature of the art tasks. Although the range of variation of obtained 'rs' values across art tasks suggests the possibility that there may be some art tasks that bring out and intensify those qualities which are identifying characteristics of both criterion measures, and these qualities are equal, but different, in their ability to identify artistic creativity, none of the art tasks employed within the context of this study proved to be of this nature.

It was shown statistically from the testing of Hypothesis II that there was a significant degree of variation among the rank-orders assigned to the subjects' products of the different art tasks. This was true of the evaluation results of each judge group. In each instance the obtained average correlation coefficients approached zero. Variation, and not consistency, characterizes the performance of the subjects in art making. The art products, as a reflection of the subjects' creative artistic performance, indicate that the art task and media plays no small part in determining the degree of "creativeness" of an individual's art work and how "spontaneously" he would handle the art materials in executing the art product. Any of the art tasks and media included in this study, or those art tasks and media which were incorporated in the development of the two criterion measures, could be questioned as to their validity as appropriate vehicles for determining creative ability. The degree of variation noted among rank-orders of the art task' products warranted the pursuance of the possible source(s) of variance which would operate systematically across art tasks.

In the testing of the secondary hypotheses, it was shown statistically that each of the variables within the hypothesized dichotomies (two- and three-dimensionality, fluid and inert art media, direct and indirect sensorial contact with the media) did not correlate when the rank-orders of the products of art tasks which were characteristic of each variate were isolated and compared. A significantly positive correlation was obtained from the evaluation

results of the Six Criterion Judges between the two rank-orders of art task's products which required direct sensorial contact with the media, but the correlation coefficient (.557) did not surpass the prescribed .70 degree of correlation required to be considered meaningful. It was also suggested that if more than two samples had been available that the 'rs' value of .557 could have been much lower and more in keeping with the obtained results from testing the other variables.

There were no correlation coefficients obtained from any art task pairing which were considered important enough to draw a meaningful conclusion. All obtained 'rs' values which achieved at least .10 level of confidence on a two-tailed test were from the evaluation results of the Six Criterion Judges. By isolating and extending the pairings between the rank-orders of art task' products which were significantly positive correlated, it was suggestive that depth of involvement with the problem or subject was the primary reason for their relatedness. Furthermore, the evidence suggested the possibility that possible hypothesized variables which could be considered as influencing "spontaneous" performance in the visual arts cannot be adequately determined until a high degree of consistency of artistic performance by the subjects is guaranteed; that is, subjects must have confidence in the resources which they employ in artistic problem-solving before their levels of "spontaneity" can be measured adequately or before affective variables can be identified.

The testing of Hypothesis VI was not completed due to the negative interjudge correlation coefficients obtained between the two teachers who composed the Teacher Judges.

CHAPTER V

SUMMARY OF PROCEDURE, FINDINGS, AND CONCLUSIONS

RESTATEMENT OF PROBLEM

This longitudinal study was made to compare two evaluative criterion scales in the scoring of art products of the same population sample. These criterion scales were purported to measure in terms of their authors' definitive and normative statements the degree of creativeness of art products.

The purposes of this study were (1) to determine whether there were significant differences between and among the rank-orders of art products of the same sample of subjects over a five month period with the rank-orders to be determined by three judge groups employing two criterion measures, and (2) to analyze any ranking fluctuations of the subjects' art products among different art tasks.

NEED FOR THE STUDY

There has been a need for an exploratory study which would entail a comparison of the more promising evaluative measures for rating art products of the same population sample over a period of time. There has been a need to determine the extent and nature of fluctuations in artistic quality, as determined by these evaluative measures, of art products of the subjects over an extended period of time. Such an extended sampling of art tasks, media, and materials

and the evaluation of these criterion scales would indicate the feasibility of the employment of these criterion scales by art teachers in evaluating their teaching effectiveness or for identifying potentially gifted art students. Furthermore, there is the need to determine if there are some variables which operate systematically in influencing the consistency of artistic performance.

BACKGROUND OF STUDY

The state of knowledge about creative behavior in the visual arts is negligible. One thing is clear: creative behavior involves complex relationships of variables which now lack clarity, and this complexity is compounded by the nebulousness of what constitutes art and artistic behavior. In spite of the risk of being premature, some researchers have isolated variables which they claim identify artistic abilities. There is little consensus of opinion among these investigators as to what variables are related to the aesthetic and creative qualities of art products. This state of affairs could lead to serious prejudgment and error.

Coupled with the question as to what constitutes the essential criteria for judging the creativeness of art products is the question of whether a single criterion is as effective as multiple criteria in making art judgments. Researchers who have devised single criterion measures suggest that their instruments are effective in rating art products in spite of increasing evidence that other variables could influence the ratings of subjects' art products in any specific art

activity or across art tasks. Whether or not a global-type criterion measure can be effectively employed in identifying and rating aesthetic and creative qualities of art products has important implications for art education. The ability of art teachers to effectively evaluate their art instruction or to identify potentially gifted art students probably depends upon the simplicity of the evaluative measures employed; the need to use multiple criteria for such evaluation could make evaluating exceedingly difficult.

Many criteria scales dealing with aesthetic and creative qualities of art products have been devised by both inductive and deductive means. Inductively, researchers have measured variables which they have decided were critical or essential to creating a work of art. They have initiated these variables from seemingly a number of sources among which are the theory and research findings in the behavioral sciences, the predominant art philosophy of the time, and their personal values derived from past experiences in art making and art teaching. Deductively, researchers have identified behavioral characteristics which are seemingly common among the working procedures of creative artists or common visual elements of art works of recognized, creative artists. Some criteria scales are used for purely descriptive purposes; that is, they are used to objectively describe and quantitatively measure the various visual aspects of art works. Other researchers have devised scales which they claim to measure the essential attributes of art; that is, particular visual characteristics of art products are valued more highly than others as being essential

to and a measure of the creative worth of art works. Some evaluative scales, although different in terminology and definitions, purport to measure the same qualities of art products. This is the case of the evaluative scales pertinent to this study. Both of these criterion measures are claimed to measure the degree of creativeness of art products, and this study was devised to compare the results of employing these measures for evaluating the art products of the same population over a five-month period of time.

SUMMARY OF THE PROCEDURE

Population and Sample

The final number of subjects selected for testing the hypotheses of this study were 22 eighth grade students enrolled in a junior high school art class. The population of this school was considered heterogeneous in terms of socio-economic and scholarship attributes as well as artistic ability and experience; however, this was not statistically determined. Furthermore, these subjects were at the end of the period of their schooling in which formalized art training had been minimal and at the beginning of the period of schooling in which art instruction from qualified art teachers was available.

Instruments Used

Two criterion scales were used for evaluating the subjects' art products. The "Typology of Creativity" for the visual arts was developed by Elliot Eisner for classifying and defining three levels

of creativity within two loci: form and subject. Although this scale is based on differential criteria, it is essentially a "global-type" measure because it measures one variable, creativeness, on a continuum. The other scale, "Six Criterion Elements of Spontaneity," was developed by Gloria Bernheim and is a modification and a simplified version of a similar instrument developed by Kenneth Beittel and R. C. Burkhart. This is a "global-type" measure whose six elements directs one to different aspects of the art product's surface for identifying the degree of spontaneity in its artistic execution.

While the Bernheim' scale requires judges to rank art products on the degree of apparent spontaneity in their artisitc execution, i.e., lack of contrived delineation of subject and content and the lack of restraint in applying the art media, the Eisner' scale directs judges to rank art products on the degree of novel elaboration and combination of their subject matter and forms or the degree of originality in the production of subject matter and form. Essentially, the Bernheim' scale is used as a measure of the quality of surface treatment of the art products, and the Eisner' scale is used as a measure of the innovative quality of the art productions' subject matter and forms and organizational relationships of the subject matter and forms.

Procedure

A photographic record of the subjects' art products was accumulated over an approximately five-month period which resulted in

eight art task samples involving various art media and problems. The products of each art task were presented for evaluation within a single format, and the "forced-ranking" procedure was used in assigning ranks to the art products.

The study involved the training of three groups of judges in the proper ranking procedures. Two three-man judge groups were formed, and each judge group was trained in the use of one of the evaluative criterion measures for judging and ranking art products. The judges designated the Typology Judges used the "Typology of Creativity" measure, and the judge group designated as the Six Criterion Judges employed the "Six Criterion Elements of Spontaneity" measure. Each judge ranked art products independently. The third judge group was composed of the two teachers involved with the subjects in a "team-teaching" situation and were designated the Teacher Judges. The Teacher Judges used both criterion measures, but separately, in judging and ranking the subjects at the end of the research period rather than judging each art task sample. This procedure was followed because of the teachers' intimate knowledge of the subjects and their art products, and the researcher's desire to have an evaluation of the subjects and not just their art products. These two judges also ranked the subjects independently. Two nonparametric statistical tests were used to determine interjudge reliability and to test the stated hypotheses. They were Kendall's Rank Correlation Coefficient: W and Spearman Rank Correlation Coefficient: r_s.

SUMMARY OF FINDINGS AND CONCLUSIONS

Judges and Judging

Kendall's Rank Correlation Coefficient: W was used as the statistical model for determining interjudge reliability. On both the training art tasks products (five art task sets) and the products of the eight art tasks which comprised the samples of the subjects under study, moderate to high interjudge correlations were achieved. There was one exception. During the training period, both the Six Criterion Judges and the Typology Judges obtained low positive interjudge correlations ('W' values of .369 and .324 respectively) from their evaluation results of the products of the non-objective art task. It was pointed out that the nature of this art task resulted in nearly all subjects producing art products which were characterized by a similar high degree of "spontaneity" and "creativeness." This probably made rank differentiation difficult on the part of the judges. Except for the difficulty the judges experienced in evaluating the products of the non-objective art task, obtained 'W' values ranging from .933 to .639 on the training set and resulted in positive significant correlations at the .001 to .05 levels of confidence. Interjudge correlations on the eight art tasks which comprised the basis of this study resulted in 'W' values which ranged from .931 to .569 and were significant at the .001 to .05 level of confidence.

Furthermore, the high interjudge correlations obtained at the very outset of the judge-training session indicated the ease with

which judges can learn to use the two criterion measures. The criterion scale, "Six Elements of Spontaneity," can generally be reliably employed by judges using the rank-order method of evaluating art products because only one "more than/less than" decision needs to be made, and this decision is not easily influenced by other variables. However, art tasks which employ art materials or processes which inherently result in free and "spontaneous-like" qualities in a large number of art products from the same sample makes rank differentiation difficult and can cause lower interjudge reliability in the employment of this measure.

In spite of the fact that the Typology Judges could reliably use the criterion scale, "Typology of Creativity," in ranking art products, there seemed to be more problems inherent with the employment of this criterion measure. In discussion with the Typology Judges after their evaluations were completed, the judges made some critical comments about the "Typology of Creativity" scale in retrospect. The single decision of "more than/less than" apparently does not always hold true when using the "Typology of Creativity" for rank-ordering art products. When two or more art products display different loci (form and subject) which appear equal in value as far as the level of creativity is concerned, a judge will apparently rely on criteria alien to this criterion measure in determining the art products' rank-order. In such decision-making, the criterion judges apparently relied on most was complexity.

A similar problem with decision-making was expressed by the Typology Judges with the ranking of art products at the lowest level

of creativity, "Boundary Pushing." The assumption had been made that some degree of novelty would be present in every art product because of the subjects' naive perception, lack of skill in handling the art materials, and the teachers' emphasis on originality. That is, no art product would be completely void of "novelty-appearing" features. The validity of this assumption must be questioned because some subjects were obviously intent upon reproducing with art materials their concept of a chosen object. When two or more subjects chose the same object to make, the final art products were generally sterile, stereotyped, and very similar. In such situations, the judges, first of all, expressed the opinion that the art products had no measurable amount of "novel elaboration," and secondly, they had difficulty in determining the rank-order of the products. Because these products identified with the lower level of creativity were apparently not differentiable, the Typology Judges apparently relied unconsciously on the degree of displayed complexity to determine the art products' rank-order.

There was a consensus of opinion among the Typology Judges that they had relied on the degree of complexity when making decisions as to the rank-order of art products which seemed to have equal value in terms of the definitions of the "Typology of Creativity" measure. No other alien criterion was suggested by the judges as being used in their decision-making. The Six Criterion Judges did not express having similar difficulties, and the problem probably did not appear with the spontaneity measure because it has a built-in complexity factor in one of its six elements.

The Typology Judges also commented upon the fact that nearly all of their evaluations were based on "Boundary Pushing"--the lowest level of creativeness on the "Typology of Creativity" measure. Naturally, this would make rank differentiating difficult. However, this low level of creativity should not be construed as being a characteristic of eighth grade students. It may have been due to the nature of the art program offered to the subjects.

An interesting and suggestive outcome was found in relation to the judging experiences of the Typology Judges. One of the Typology Judges commented that he probably evaluated art products differently because of his training and interest in sculpture. He was especially concerned with his evaluation within the locus form. Furthermore, it was noted of the paired judge correlations across art tasks that he and another Typology Judge, who had an interest in and taught ceramics and ceramic sculpture, obtained the highest paired correlations among the paired correlations of the three judges on only two art tasks: ceramics and sculpture. Of the three art tasks classified as three-dimensional, only these two resulted in art products which were in full relief. The possibility was suggested that one's background of training and interest in art affects his perceptual-ordering and emphasis. If this were true, then the ranking of a particular art product could be materially affected by judge training, experience, and interest. For instance, Eisner gave an example of "Boundary Breaking" as making a mosaic on a sphere. The question which could be asked is whether a judge whose training, experience, and interest

were in three-dimensional art making would consider this example as creative as a painter who would be more perceptually stereotyped toward two-dimensional surfaces.

Hypotheses

The findings of this portion of the investigation are presented in relation to the stated hypotheses.

The hypotheses were formed to test three basic questions. Of the two primary hypotheses, Hypothesis I was formed to test the question of the degree of correlation between the employment of the two criterion measures in identifying the creativeness of art products of the same population of subjects. The testing of Hypothesis I attended to the degree of correlation on each art task and then the across art task average. Hypothesis II, the second primary hypothesis, attends to the consistency of creative performance of the subjects as reflected by the ranking of their art products among the art tasks on the basis of each criterion measure.

Hypotheses III, IV, and V are the secondary hypotheses and they were formed to test the question of whether there are particular systematic causes of variance of creative performance in art making. The variables analyzed through the testing of the secondary hypotheses were space dimensionality, resistance of art materials, and sensorial empathy with art materials. The variables served as a basis for dichotomizing art task products in order to determine, first of all, if there was a significant variation of rank-orders between the two

categories of art tasks which were defined by the dichotomy; and secondly, to determine the degree of correlation among the rank-orders of products of the art tasks which comprised each of the categories of a dichotomy. Any meaningfully significant correlation between the rank-orders of products of art tasks within any one or more categories of the three dichotomies would have lent evidence to the hypothesis that there are variables which operate systematically in determining the consistency of creative artistic performance as defined by the two criterion measures.

Space dimensionality as a source of systematic variance was tested by Hypothesis III, and the art tasks and products were dichotomized into two- and three-dimensional categories for this purpose. Hypothesis IV was formed to determine if the level of resistance to modification of different art media would act as a source of systematic variance in "creative" and "spontaneous" performance of the subjects under study. For this purpose, art tasks and products were dichotomized by the inert and fluid characteristics of the art media employed. To test Hypothesis V, which was formed to determine the degree that sensorial empathy with the art media operated as a systematic variable, the art tasks and products were dichotomized into indirect and direct sensorial contact with the art media categories.

Hypothesis VI was not completely tested because of the negative interjudge correlations obtained between the two teachers who formed the Teacher Judge group. The question underlying this hypothesis was

whether the teachers with an intimate knowledge of the subjects as art students could, without prior training in the use of the criterion measures, arrive at rank-orders of the students' "creativeness" and "spontaneity" which would be similar to the ranking of subjects, as reflected by their art products, by the Typology Judges (creativity) and the Six Criterion Judges (spontaneity).

For testing the hypotheses, the Spearman Rank Correlation Coefficient: rs was used throughout. This statistic, sometimes called rho, was represented by 'rs.' Direction of difference was not predicted in any hypothesis tested and this required a two-tailed test. All 'rs' values which were significant at the .10 level of confidence were noted; however, the .70 degree of correlation had to be achieved before any 'rs' value could be confidently considered useful or having real predictive strength. N (number of numerical ranks) varied according to the number of art products in rank-orders being compared.

Hypothesis I: There will be a statistically significant difference between the pooled rankings of the Typology Judges and the Six Criterion Judges on each art task and across art tasks.

The obtained 'rs_{av}' value of .402 between the two judge groups' rankings across the eight art tasks was neither meaningful nor significant. This result lends evidence that the two criterion measures do not identify similar behavioral characteristics. The range of 'rs' values (-.033 to .673) obtained between the two judge groups on the rank-ordering of the products of the eight art tasks indicated

that a different series of art tasks could produce different results. It needs to be borne in mind that this was not a random sample from all possible kinds of art tasks which are presented to junior high school students. For instance, if the cartoon art task ('rs' value -.033) was eliminated from this art task series, the 'rs_{av}' value would increase to .464 which is significant at the .05 level of confidence, although .464 is still well short of the prescribed .70 degree of correlation needed to have been considered important. On the other hand, if most of the art tasks were dominantly three-dimensional problem solving tasks or at least evenly divided between two- and three-dimensional problems, the 'rs_{av}' would have probably been much lower (the 'rs' values obtained on the ceramics and sculpture art tasks were .306 and .278 respectively).

The variation of obtained 'rs' values across art tasks suggest the possibility that there may be some art tasks that bring out and intensify those qualities which are identifying characteristics of both criterion measures, and these qualities are equal, but different, in their ability to identify artistic creativity. Although the results of testing Hypothesis I do not prove or disprove such a possibility, none of the art tasks as presented to the subjects in this study resulted in a considered important correlation between the rank-orders of their art products.

It is interesting that the correlation coefficients obtained between the two judge groups on each art task could be reasonably justified, although not statistically; that is, by knowing the nature

of the task, how it was presented, how the subjects responded to it, and analyzing the final products the obtained 'rs' values were justifiable. An example was given previously of the judging outcome on the rank-ordering of the products of the fantasy drawing task which resulted in a 'rs' value of .673. This correlation coefficient is close to the prescribed .70 degree of correlation needed to be considered important and in a replication of this study the 'rs' value could possibly surpass the prescribed .70 degree of association.

The fantasy drawing task "freed" the subjects from reality to some extent; that is, the subjects did not do genre pictures but many of them relied on grotesque figures derived from such illustrative literature as Mad Magazine. The final products were generally marked by a "spontaneous" quality not exhibited in their previous work and, at the same time, "creative" to the extent that the products were novel in form and subject. It is questionable, however, whether there is a meaningful correlation between the manifestations of fantasy and any of the multiple definitions of creativity or imagination. The fantasy drawing task may have value in achieving particular objectives of art learning behavior, however, it is doubtful that the products of the fantasy art task would reflect the insight and imaginative interpretation of a subject's personal, social, and physical environment which is so much a part of recognized creative artistic efforts. What the products of this task might do is reflect a subject's psychological distance from reality.

The reality which needs to be reemphasized is the results of testing Hypothesis I, and that is, there were no 'rs' values obtained

between the rank-orders of the two judge groups on any art task or across art tasks which could be considered important.

Hypothesis II: There will be a statistically significant difference among the pooled rankings assigned to the products of different art tasks by the Typology Judges and the Six Criterion Judges.

In comparison to Hypothesis I in which the problem was to determine whether the two judge groups would arrive at similar rank-orders on each art task and across art tasks, the question underlying Hypothesis II was to determine whether each of the two judge groups would arrive at similar rank-orders among art tasks.

Statistically, there was a significant degree of variation among the rank-orders assigned to the subjects' products of the different art tasks. The obtained 'rs_{av}' values for each judge group approached zero relationships (obtained 'rs_{av}' values were .035 for the Typology Judges and .031 for the Six Criterion Judges).

The results obtained from testing this hypothesis clearly indicated that variation and not consistency characterized the performance of the subjects as reflected by their art products. This is not to say that there were no subjects who did not display consistent "spontaneous" or "creative" performance across art tasks because change of rank does not necessarily mean change of performance; however, the degree of change of rank-orders among art tasks did indicate that there was a good deal of fluctuation of art-making performance as measured by the two criterion measures.

Doubt is cast upon the single art task argument for identifying creative or artistic talent. Any of the art tasks and media included

in this study or those art tasks and media which were incorporated in the development of the two criterion measures could be questioned as to their validity as appropriate vehicles for determining creative ability. For example, the use of crayon as being the most appropriate medium for executing the product of any art task test has its limitations. The specification of a particular art medium because it is the most familiar medium to the subjects disregards the quality of the experiences that the subjects have had with the medium and their personal reactions to it.

In sum, any of the art tasks and media included in this study, or those art tasks and media which were employed in the development of the two criterion measures, could be questioned as to their validity as appropriate vehicles for determining creative ability. The analysis of obtained results from testing the secondary hypotheses cast further doubt on the assumption that there are certain art tasks and media which are best suited for visually soliciting one's creative potential.

Secondary Hypotheses

The secondary hypotheses included Hypothesis III, IV, and V and were inductively devised to test possible contributing systematic sources of variance. They were devised on the basis of variables whose dichotomies would separate art tasks as to spatial problem-solving strategies, limitations of art materials, and directness of sensory contact with the media.

The results from testing the secondary hypotheses were very similar, and for the purpose of this summary they will be reviewed together.

Hypothesis III: There will be a statistically significant difference between the pooled rankings of the subjects' two- and three-dimensional art products. This difference will be found between the pooled rankings of each judge group.

Hypothesis IV: There will be a statistically significant difference between the pooled rankings of the subjects' art products executed with fluid art media and those executed with inert art media. This difference will be found between the pooled rankings of each judge group.

Hypothesis V: There will be a statistically significant difference between the pooled rankings of the subjects' art products which entailed direct sensorial contact with the art media and those products which entailed indirect sensorial contact with the art media. This difference will be found between the pooled rankings of each group.

The evaluation results from each judge group clearly accepted that there was a statistically significant difference between the rank-orders within the hypothesized dichotomies. However, to say that the variables under study in each of the secondary hypotheses operate as systematic sources of variance was considered erroneous as other variables were obviously operating. This conclusion was made because in isolating those art tasks which were identified with either of the two categories within each dichotomy and extending the rank-order pairings among these art tasks, it was discovered that there was a minimal degree of relatedness among the rank-orders of the products of these art tasks.

There was one exception. In computing the correlations between the rank-orders of art tasks which required direct sensory contact with

the art media (ceramics and papier mâché plaques), a significantly positive correlation was obtained from the evaluation results of the Six Criterion Judges ('rs' value of .557 which is significant at the .05 level of confidence); however, the 'rs' value did not surpass the prescribed .70 degree of association required to be considered important. A negative correlation was obtained from the evaluation results of the Typology Judges ('rs' value -.191). It was considered doubtful that direct sensorial contact with the art media would operate as any more of a systematic source of variance on "spontaneous" performance in art making than any of the other hypothesized variables. To justify this contention, other variables which seemed to be operating in the making of ceramics and papier mâché plaques were studied.

It was noted that all significantly positive correlations were obtained from the evaluation results of the Six Criterion Judges. From the evaluation results of the Six Criterion Judges, twenty-eight possible correlations were computed between rank-order pairings among the eight art tasks. Three of these correlations reached the .10 level of confidence or higher, although no 'rs' value surpassed the prescribed .70 degree of association. Besides the correlation noted between the rank-orders of products from the ceramic and papier mâché art tasks, significantly positive correlations were obtained between the paired rank-orders of the products of the ceramic designs and ceramic art tasks and the ceramics and the papier mâché plaque designs. The 'rs' value obtained between the rank-orders of products of ceramic designs and ceramic art tasks was .536 which was significant at the

.05 level of confidence. The 'rs' value obtained between the rank-orders of the ceramic and papier mâché plaques design art tasks was .557 which was significant at the .10 level of confidence.

By selecting out these four art tasks and extending the rank-order pairings among them some suggestive implications were noted: (1) ceramic design and papier mâché plaque design art tasks were colored comprehensives for the ceramic and papier mâché plaque art tasks, (2) ceramic design and papier mâché plaque design art tasks were executed with crayon, (3) ceramic and papier mâché plaque art tasks were three-dimensional problem-solving tasks, and (4) the products of the ceramic and papier mâché plaque art tasks, besides requiring direct sensorial contact with the art media, were the final products of fairly extensive preplanning which was the result of carrying through from the color comprehensives noted in (1) above.

From the analysis of the extended rank-order pairings of these four art tasks, it appeared that depth of involvement with the subject or presented problem was the dominant contributing factor which influenced the subjects' consistency of "spontaneous" performance in art making; however, it may be that two or more variables such as direct sensorial contact with the art media and depth of exploration of the problem would determine the consistency of a subject's "spontaneous" performance in art making.

The same explanation is hardly valid in considering the 'rs' values obtained from the judging results of the Typology Judges. It was shown from the testing of Hypotheses III, IV, and V that no art

task pairing achieved the .10 level of confidence and only one 'rs' value reached the .20 level of confidence.

Some of the lower correlations from the evaluation results of the Typology Judges could have been due to lower interjudge correlations, but it is also plausible that the difference between the defining characteristics of the two criterion measures could be a major factor. It may be that one's level of "spontaneity" in handling the art materials would be influenced less by unfamiliarity with the medium, subject, or kind of problem than would one's level of "creativeness" with form and subject.

For one thing, unfamiliarity of the inherent characteristics of a medium can obliterate creative ideas. Too often novel, inventive concepts of form and subject are covered over because of the lack of skill in handling the medium or there is a lack of compatibility between the medium and the delineated symbol.

In other instances, the subjects seemed to be concerned with directing their energies toward getting-to-know the subject and not in the novel elaboration of the subject or in exploiting its aesthetic possibilities.

These varied orientations by the subject toward the various art tasks could be a major source of variance in the obtained correlation coefficients among art task products evaluated on the basis of both criterion measures, but most particularly, the "Typology of Creativity" measure. It is well to reiterate that, whereas the judging of the "spontaneous" level of an art product is mainly based

on surface treatment, the judging of the level of "creativity" is mainly oriented toward the basic structure of an art work, and this underlying characteristic can be hidden by lack of skill in executing the art work in an unfamiliar medium.

No systematic source of variance was identified in the testing of the secondary hypotheses. It was apparent from the range of variation among product ranks of the art tasks that the variables under study in each of the hypotheses would not operate in the art productions of subjects of a population sample similar to the sample used in this study; however, it is not accepted that the hypothesized variables tested in this study would not operate systematically in the art productions of a differently defined population or under different conditions.

Although depth of involvement with the problem needs to be more thoroughly tested, it was suggested that a greater degree of consistency of art performance must be achieved by subjects under study before possible systematic variables which would determine types of artistic behavior could be adequately tested.

CHAPTER VI

DISCUSSION AND EXTENSIONS

One of the purposes of this study was to compare the results of employing two criterion measures in the evaluation of art products. Evidence was obtained which indicated that the two criterion measures, "Six Elements of Spontaneity" and the "Typology of Creativity," do not agree in the value and order of the visual qualities of the products of art tasks analyzed in this study. The testing of the hypotheses resulted in a large degree of variance of correlation coefficients from paired rank-orders among art tasks. No evidence was discovered, nor was such evidence sought, which would indicate which of the two criterion measures would more adequately measure the degree of creativeness of art products. Nor was there any evidence sought or discovered which would prove that the criterion measures could actually identify and measure the creativeness of art products. For information on the normativeness of the two criterion measures, the writer had to rely on the written reports by the authors of the criterion measures. Perusal of these reports raises a number of questions concerning these two criterion measures.

With rare exception, art scales devised in recent years have been neither normalized nor standardized; that is, they have been developed on the basis of a small sample of subjects or from samples of subjects with similar parameter characteristics. This was true

of the two criterion measures used in this study; however, normalization or standardization of the two measures is not as important as their applicability to samples of subjects with different parameter characteristics.

Eisner, in developing his "Typology of Creativity," used as his sample of subjects eighty-five boys and girls attending a Midwestern private elementary school. Further information as to other characteristics of this sample were not reported. There are no reports that this measure has been used on other samples with different parameters in order to obtain other distributions or statistical results for comparison.

In the development of the "Spontaneity Domain" which culminated with Bernheim's "Six Elements of Spontaneity," the samples of subjects used were college students and high school juniors. Other information about the samples used was that junior level art education majors were used in the college sample. There is no indication in art education literature that other samples of art products of subjects with different parameter characteristics were investigated in relation to this measure.

In respect to the applicability of the two criterion scales to different population samples, the reliability and validity of these criterion scales must be considered. Two sources as to the reliability and validity of these criterion measures used in this study can be

drawn upon: (1) written reports by the authors of these measures, and (2) the results from the present study.

Considering reliability of the two criterion measures, first of all, the evaluation of the reliability of any measure reduces to a determination of how much of the variation between rank-orders is due to systematic differences among the individuals in the group and how much to inaccuracies in measurement of the particular individuals. Generally, systematic differences and inaccuracies of the measure are not investigated if the judges are in significant agreement in the employment of the measure. However, in this study, the two criterion measures were being compared and particular findings were discovered which makes the discussion of the reliability and validity of the two measures desirable. In the case of the two criterion measures used in this study, the first index to their utility was the extent of interjudge agreement. Eisner reported in his study that interjudge agreement was moderately achieved. There was no report on either judge characteristics or interjudge agreement from any of the investigators of the "Spontaneity Domain." In the present study moderately high interjudge agreement was achieved on both evaluative scales at the very outset of the judge training period. It was concluded that with a short training period judges could reliably employ the two evaluative scales in the ranking of art products.

Although systematic differences between judges were not sought, the possibility of such differences existing was discovered in post-study

discussion with the Typology Judges. An interesting, but inconclusive, outcome of this discussion with these judges was the possibility that a judge's interest and training in art could influence his perceptual ordering in the use of the "Typology of Creativity" measure. If it were simply a "yes or no" decision a judge is called upon to make, it would probably make little difference; but in instances in which "more than/less than" decisions are made on the basis of a particular criterion or criteria a disparity of measurement between two judges of different art training or interest may result. For example, if a sculptor and a painter were to judge art works on a composite of form and color qualities, it is possible that the sculptor would tend to emphasize form, quite unconsciously, in his decision. Conversely, the painter may tend to attach more importance to color than form. Previously, an example of "Boundary-Breaking" creativity was given to be a student who makes his mosaic upon a sphere rather than a flat surface. The question asked was whether a sculptor, with his experience with decorating and enhancing a multitude of various topographical surfaces, would perceive this example as novel as would a painter who is more perceptually stereotyped toward two-dimensional surfaces.

Differing perceptual ordering in attending to an object or situation is not a surprising or new possibility. In the selection of art judges, however, this writer is not aware that this aspect has been a consideration. There is probably an unanimity of opinion among investigators of perception that interests, attitudes, temperament,

needs, and aptitudes affect how one perceives. Klein draws the conclusion that drawing preference is influenced by one's training.[1] Fritzky points out that people are "differentially sensitive" to the various dimensions of design and goes on to say:

> If, then, they (two people) experience the composite, ABCD, it is merely a manifestation of the degree to which each of them can differentiate certain aspects of the stimulus dimensions and the perceptions are the same for both--they simply are responding to different "versions" of ABCD.[2]

It would seem logical, then, that any consideration intended to aid in the formulation of a judgment decision will usually reflect some "hidden agenda" on the judge's part. It is also possible sometimes that the judge may not be conscious that this is true, but its danger to a "good" decision is still present.

No matter how "criteria-centered" a judge may be, he is bound to stress at least a little more enthusiastically a solution that will cause him the least trouble, especially if it is a good decision though not the best one possible.

It should not be overlooked that teachers and children are also judges. Teachers are continually making judgments of their program

---

[1]Stephen P. Klein, A Description of Points of View in Esthetic Judgements in Terms of Similarity Dimensions, Princeton, N.J.: Educational Testing Service, 1967 (RB-67-53).

[2]Ferdinand Fritzky, Aesthetic Preference for Abstract Designs as a Function of their Perceived Complexity, Princeton, N.J.: Educational Testing Service, 1963, (RB-63-27), p. 35.

and their students. Students are continually making judgments during the process of art making. It has been the writer's experience that evaluation is the weakest link in the teaching act. Apparently it is a lack of understanding of the evaluation process and/or the inability or reluctance to make objective evaluations which are apparently the reasons for poor evaluation practices. It is claimed that children must learn to be self-critical if realistic self-concepts are to be developed; yet, many teachers claim the children are too young to evaluate their own work. More than likely the teachers do not know how to develop evaluation ability within their charges. Very little research has been done in art education into the evaluation process as it relates to the teaching and learning act, and there is a real need to do so.

It is the contention of this writer that no matter what criteria are used for evaluating the artistic or creative qualities of children's art production, they will be of little value unless they can be effectively employed by the classroom teacher. Whether it is for the identification of certain artistic abilities or to determine the effectiveness of an art program, it is the teacher who has the "facts," and these facts are at times important mitigating considerations in a particular situation. Little confidence can be put in the utility of the two criterion measures used in this study from the negative inter-judge correlations obtained from the two teachers who composed the Teacher Judge group. To what extent mitigating circumstances played

a part in the negative interjudge correlations was not determined. Whether differences were due to some systematic variable or whether it was a chance occurrence was not investigated.

In any case, the utility of art scales by the classroom teacher should be a continuing important consideration of researchers in their searches and development of essential criteria of creative art production. One study which offers some optimism that classroom teachers can make reliable judgments of their students' artistic creative ability was conducted by Alford.[3] Two hundred and eighty-five subjects previously judged by their teachers to be either high, medium, or low in artistic creative ability were administered four tests of art judgment, art performance, general creativity, and ingenuity. The teacher judgments of the students were sustained by performance of the students on the tests and also by an art contest which was independently juried.

There is a need to investigate more thoroughly personality characteristics of art teachers especially in situations which require the teachers' cooperative efforts. With the trend toward team-teaching and open-area instruction, this may become very critical for art education. Assuming that art teachers are creative, then they will

---

[3]Mary Lee Alford, Teacher Judgments as Related to Certain Predictors of Artistic Creativity in Senior High School Students, North Texas State University, 1964 (unpublished doctoral dissertation).

possess those characteristics of a creative person described in the literature on creativity. This would mean that they are imaginative, innovative, flexible, and ideationally fluent in the pursuance of solving problems which result from personal needs and interests. However, literature describing creative personality characteristics is also replete with such descriptive terms as independence of thought and action, autonomy, resistance to enculturation, self-differentiation, and self-actualization. A mundane summary of these latter creative characteristics is that a creative person is biased and opinionated, hardly a desirable person for cooperative efforts in teaching.

Besides the possible systematic differences which may determine the utility of the two criterion measures, there is apparently some inaccuracy in using these criterion measures in the evaluation of particular kinds of art products. Many of the possible inaccuracies of measurement are due to the inability of the defining characteristics of particular classifications and elements to distinguish between art products with similar values of the defined characteristics.

An example of experiencing this kind of difficulty in evaluation occurred during the training period for judges in the use of the criterion measures. The products of the judge training set of art tasks which were non-objective in nature resulted in low, nonsignificant interjudge correlation coefficients for both judge groups. In relation to the definitions contained within the "Typology of Creativity," all

of these art products were classified as "Boundary-Breaking" or by definition they were all "utterly new" in form and subject. The question is that once a number of art products are identified as "utterly new," how does one decide if one art product is more or less "utterly new" in comparison to another art product? Assigned scores rather than numerical ranks would not solve the problem. Even if there were criteria which would aid in ranking the "utterly new," difficulty would be encountered because of the degree of similarity between so many art products. This was also apparently the problem of the judges who employed the spontaneity measure. Although the "Six Criterion Elements of Spontaneity" had six different elements which the judges could use in making "more than/less than" decisions, the art products of the non-objective art tasks were all of high value on each of the six elements.

In reference to the spontaneity measure, one factor which appears to play a dominant role in determining the degree of spontaneity exhibited in art products is the art materials and processes used. Some art materials, because of their intrinsic qualities, require a more loose manipulation in order to achieve "satisfactory" results by the manipulators. Coupled with the nature of particular assigned art tasks, the products would have an inherently "spontaneous-like" quality. The result would be that judging results of these kind of art products may not be a "true" measure of the subjects' spontaneity. Furthermore,

such a possibility points out that judging results could be influenced with or without intent.

An example in which biased results would be most detrimental would be in the selection of students from multiple schools for a class for gifted art children. Any one teacher with intent or with a preference for particular teaching materials and processes which inherently result in spontaneous appearing art products could have a predominant, and not necessarily the most artistically gifted, number of his students selected for this special art class.

Such a misrepresentive group could also be selected by judgments on the basis of the "Typology of Creativity." Even if no criteria scale was used at all and the selection was determined on the basis of a jury of "art experts" with their individual likes and dislikes, the outcome could be similar. It is quite easy to become satiated with the usual offerings of junior high school students and be attracted to the unusual, even if the unusualness is task-determined, accidental, naive, or attributable to the nature of the art materials and processes rather than an outcome of any creative or imaginative intent of the students.

The Typology Judges encountered three other difficulties in employing "The Typology of Creativity" measure in the evaluation of particular kinds of art products. An example of one kind of difficulty would be in the case where two art products would be identified as being at the same level of creativity but on the basis of different loci,

subject and form. Consider two art products at the same level of creativeness but displayed through different loci. If each art product displays the same degree of the particular locus's defining characteristic, which takes precedence in ranking? This kind of difficulty is not necessarily a weakness of this scale as the scale was originally designed for assigning raw scores on an interval scale. This difficulty could have been overcome if a score had been assigned instead of a numerical rank; then at least ties would have been allowed. This could have been done and still have made the data amenable to nonparametric statistical models. As it were, the rank-order of measurement employed in this study had no such allowance. The Typology Judges suggested, in retrospect, that they may have unconsciously shifted to the degree of complexity in making a decision in this situation.

The second difficulty the Typology Judges apparently had and one which suggests that the reliability of utilization of the "Typology of Creativity" may be less than desirable, was the difficulty of determining numerical ranks when a large number of art products tended to fall within one level of creativity. It was the general consensus of the judges that nearly all of the products from any one art task fell within the "Boundary-Pushing" level of creativity. This is the lowest level of creativity in the typology. Furthermore, most art products were characerized by novel elaboration of subject with very few reflecting novel elaboration of form or novel combination of form or subject. This made ranking difficult, and in this case the assigning

of scores would have worsened the situation because the number of tied scores would have been excessive which would have made rank-order comparisons meaningless. Again, the judges related that they may have unconsciously shifted to the degree of complexity in making their decisions. No other criterion was suggested as a possible determiner in their decision-making.

The rank-order method of measurement apparently imposed a third difficulty for the judges. It had been assumed by this investigator that at least one of the characteristics defined in the "Typology of Creativity" would always be present to some degree in the art products evaluated. The bases for this assumption were the general emphasis of qualified art teachers on originality of art results and that the naivete and lack of art skills of the young artists would produce art products which would appear to some degree unique and novel. Although the assumption may still be basically sound, there were instances in which the subjects produced art products that seemed to reflect an orientation toward imitation (not copying) rather than expression. The visual outcome was often aesthetically sterile, stereotyped, and similar in conception. The judges felt that defining characteristics of even the lowest level of "creativity" were non-existent in these products, and when two art products were quite similar in conception the judges apparently relied on the degree of complexity in making their decisions.

This shift to the degree of complexity to determine numerical rank in difficult decision-making is interesting especially if it were done unconsciously as suggested by the Typology Judges. If the judges did make this shift in criteria for evaluating, they must have done it with some unison and consistency because of the obtained inter-judge agreement.

It may be well to explore complexity more fully as a possible global criterion for identifying creative children in art. Literature and research reports on characteristics of creative people reflect the complexity of the creative personality.[4] Theodore Shaw makes quite a good case for complexity as a value determiner of works of art.[5] Throughout his writing he stresses that it is the degree of complexity of an art product which makes or breaks it--that which determines if it lives on as a major work of art or sinks into oblivion. Of course, care needs to be taken in comparing great works of art with children's work; however, it is the behavior which underlies the creation of visual arts with which art education is concerned. A visual complexity

---

[4]A. H. Maslow, Motivation and Personality, New York: Harper and Brothers, 1954. See Chapter XII (pp. 199-234) in which the author discusses self-actualizing people (a universal characteristic of self-actualizing people) and draws a comprehensive verbal picture of the complexity of a creative personality as this author sees him.

[5]Theodore Shaw, Hypocrisy About Art: And What You Don't Gain by It, Boston: Stuart Publications, 1962.

scale could be devised along the lines of Eysenck's "K" Factor (asymmetry) and/or the Barron-Welsh Art Scale.[6,7] The visual complexity scale could be validated against standardized personality tests.

If one's potential creative artistic ability could be determined by the complexity of his art products, assuming that the complexity of art products reflects the complexity of the doer, then a complexity scale may be a better identifier of potential creative artistic talent than the two criterion measures employed in this study. At least complexity may not be as variable in one's art performance as his creative or spontaneity levels due to the vagaries of art task, medium and processes.

Up to this point, the discussion of the reliability of the two criterion measures has been in relation to the utility of the instruments by the judges and of possible sources of variance in making judgments on the basis of these instruments. The discussion of reliability will now be concerned with possible sources of systematic variance due to differences among the subjects under study as reflected by their art products. Primary concern will be with art materials,

---

[6]H. J. Eysenck, Dimensions of Personality, London: Rutledge, Kegan Paul, 1947.

[7]F. Barron and B. S. Welsh, "Artistic Preference as a Factor in Personality Style: Its Measurement by a Figure Preference Test," Journal of Psychology, 33:199-203, 1952.

processes, tasks, and performance as operating as sytematic differences among the individuals in the group.

The secondary hypotheses were devised to determine if such variables as space dimensionality, resistance of materials, and sensorial empathy with art materials were the cause of the variance of the subjects' "creative" and "spontaneity" levels among art tasks. Although these hypothesized systematic variables did not prove out statistically, the analysis of statistically significant correlations among art task pairings led to the discovery of some related characteristics which seemed to have implications for achieving and maintaining consistent artistic performance in art making.

Before considering these art tasks and their related characteristics, it needs to be stressed that those defined variables of the secondary hypotheses, two- and three-dimensional art problems, fluidity/inertness of art materials, and sensorial/indirect contact with the medium, did not contribute systematically to the observed variance across and among art tasks. The testing of the secondary hypotheses confirmed that there was variance between those tasks which were dichotomized on the basis of these variables. The problem was that further investigation showed that there was variance among those art tasks which were grouped on either side of the dichotomies; for example, there was a variance between rank-orders of products of art tasks which were defined as three-dimensional, those which employed a fluid

art medium, and those art tasks which required indirect sensory contact with the medium. There was one exception. From the evaluation results of the Six Criterion Judges, the obtained correlation coefficient between the rank-orders of art tasks which required direct sensorial contact with the media (ceramics and papier mâché plaques) was .557 which is significant at the .05 level of confidence. Although this correlation was particularly attractive to this investigator, there were two reasons for not putting much confidence in this finding. One, the size of the correlation coefficient was not so large as to warrant any real confidence in its predictive strength, and secondly, whereas the rank-order pairings among the other dichotomies were based on three or more art tasks, only two art tasks required direct sensory contact with the art media and the obtained correlation could have been a chance occurrence. The obvious conclusion was that other variables were in operation.

It still seems reasonable that these variables could be playing a significant role in determining the development and consistency of creative, artistic behavior. It still seems reasonable that an individual because of his unique personality would be attracted to and have greater empathy with certain art tasks, media, and processes. This would be important to know for developing one's potential for creative artistic behavior, and it would be important for realistic measurement of one's creative artistic development.

A review of research indicates there is a paucity of this kind of investigation. Silverman made one such study by comparing the effects of two- and three-dimensional art activities on art behavior. He concluded:

> In all probability, there are methods and media which will be instrumentally valuable in the development of particular abilities; the relative consummatory value may be negligible. Conversely, some activities may prove to be of consummatory value but have little affect in changing those aspects of behavior which art educators are striving to develop.[8]

This is not much of an informative conclusion but it is a beginning. Certainly many more possibilities exist as a basis for similar studies.

Although there were no correlation coefficients that were large enough to elicit any real confidence in their predictive strength or to serve as a basis for arriving at really confident conclusions, there were instances when the correlation coefficients obtained between the rank-orders of art task pairings were high enough, or nearly so, to be statistically significant at the .10 level of confidence. Some interesting characteristics were noted which were common to two or more of these tasks and not between other art tasks initiated during the period of this study. Furthermore, a fact which must be kept in mind is that these discovered relationships appeared only in the

---

[8]Ronald H. Silverman, Comparing the Effects of Two- and Three-Dimensional Art Activity upon Spatial Visualization, Aesthetic Judgment, and Art Interest, Unpublished Doctoral Dissertation, Stanford University, 1962, p. 75.

evaluation results of the Six Criterion Judges who employed the "Six Criterion Elements of Spontaneity" scale. The following discussion and conclusions will be related to the development and maintaining of spontaneity level. In spite of this investigator's belief that the following discussion and conclusions hold true for creativity as defined in the "Typology of Creativity" scale, the fact remains that similar findings as observed in the evaluations of the Six Criterion Judges were not achieved by the Typology Judges.

The obtained correlation coefficients from pairing rank-orders of products among art tasks produced four art tasks (ceramic designs, ceramics, papier mâché plaque designs, and papier mâché plaques) that were involved with paired rank-orders which were significantly similar. By isolating these rank-orders and extending rank-order pairings among them, similar characteristics were identified which two or more of the involved art tasks had in common.

One of the common characteristics noted of two art tasks whose ranks were not significantly similar was the medium used. The medium, crayola, had previously been used extensively by the subjects. This does not mean that familiarity with the medium is not a determining variable when combined with other variables. It is well to bear in mind, too, that quantity of experience tells us nothing of the quality of these experiences.

Another characteristic was design; however, there is a real reluctance to draw conclusions or even suggest implications of design

as a systematic variable. The subjects had had very little experience with designing, and in conjunction with particular art tasks it was more "planning for" than designing for some end product. Design as visual organizing was offered individually and haphazardly; hence, little faith can be put into any statement that design alone is or is not an assurance of developing and maintaining spontaneity. Not only was there less experience with design than there had been with the medium crayola, but the quality of these experiences was much more questionable.

The significantly positive correlation between the rank-orders of products of the ceramic and papier mâché plaques produced the third common characteristic. The consideration of direct sensorial contact with the art media as a source of systematic variance of "spontaneous" performance in art making has been discussed previously in relation to the degree of confidence which could be put on the obtained degree of correlation between these two art tasks. To reiterate, little confidence was elicited by the obtained .557 correlation coefficient; however, it was large enough to recommend further studies incorporating this variable.

The one common characteristic of the art products of art tasks which resulted in significantly positive rank correlations and the one which came through quite definitely was the depth of study of the chosen subject or problem for an art production. This is a reasonable finding and an approach to artistry which has been stressed

for centuries. However, in art education breadth rather than depth has been stressed for a number of years and only recently has there been a suggestion that we may be moving back toward at least some degree of depth in teaching art.

A study of particular interest and appropriateness to this study was conducted by Beittel and Mattil.[9] It is particularly relevant because one of the authors, Kenneth Beittel, was a co-developer of the original spontaneity measure, and, furthermore, the findings of the study are related to the development of spontaneity. This depth study was media oriented rather than object or problem-solving oriented.

The authors used three groups of ninth grade students. There was the usual control group who underwent a course of study which was exactly like the previous year's. It is interesting to note that the authors characterized this course of study as being basically breadth-oriented. A second group was taught according to a prescribed breadth course of study developed for the experiment. The third group, the "Depth-Group," was also taught by a prescribed course outline. Painting was the "depth" activity. It included a variety of painting media, a

---

[9]Kenneth R. Beittel and Edward L. Mattel, "The Effect of a 'Depth' vs. a 'Breadth' Method of Art Instruction at the Ninth Grade Level," Readings in Art Education, Elliot W. Eisner and David W. Ecker (eds.), Waltham, Mass.: Blaisdell Publishing Co., 1966, pp. 296-358.

study of the history of painting, and an acquaintance with contemporary painters.

Some of the findings are interesting and relate well to the present study. First, the authors found that the "Depth Group" gained significantly in spontaneity while the "Breadth Group" lost significantly. Secondly, the intercorrelates in the "Depth Group" suggest that this form of instruction will allow for a wider range of differences to be reflected by the art product than when instruction aims for breadth. As the authors point out, it appears that depth allows for more of the whole person to be related to his art activities than does breadth. If depth of study allows for a wider range of differences, then this may have allowed the judges in this study to rank art products more easily and contributed to the obtained significantly positive correlation between product rankings of art tasks which were characterized by a depth approach.

Although they are not particularly related to the present discussion, the anecdotal data collected by the authors is of interest. Students indicated that they preferred a variety of experiences, or the breadth method. However, as the authors pointed out, the study suggested that it may be well to begin earlier with children in engaging in sustained, long-term projects of depth with less yielding to their demands for variety. There is evidence in this and other studies by the authors that suggests that some kind of activities that students appear to want and are insistent upon have little learning

value for them.

There is another study which relates to depth of study of subject matter.[10] MacGregor used sixty-three students comprising four high school classes who were assigned the following treatments over a two week period: Group I was restricted to working with one arrangement of subject matter and one medium; Group II had a range of subject matter to choose from but only one medium with which to work; Group III was restricted to one arrangement of subject matter but could employ a variety of media; and Group IV had a range of both subject matter and media choice. Neither Group II (restricted in media but not in subject matter) nor Group IV (unrestricted in both media and subject matter) had any consistent advantage over the other, but both achieved at a higher level than did Group III (restricted in subject matter but not in media choice) on most of the criteria specified in the study. The most noticeable feature arising from the results was the consistency of superior achievement by Group I, the group which was restricted in both media choice and subject matter.

It is somewhat puzzling that Group III did not do at least as well as Group IV. Such a finding is inconsistent with the other results. However, it is particularly interesting in the light of the treatment period of only two weeks that the group which was restricted to working

---

[10]Ronald N. MacGregor, "Imposed Controls in Subject Matter and Art Media Choice," Alberta Journal of Educational Research, 20:103-110, No. 2, June, 1967.

with one arrangement of subject matter and one medium displayed such consistency of superior achievement. Further studies of this nature would be well worth the effort, especially if the treatments were of longer duration and incorporated different subject matter and/or problems.

Both of these studies suggest in their description something of the quality of treatments, but basically they are quantitatively described. Naturally, it is quite difficult to convey verbally such qualities; however, it is well to keep in mind that quality of instruction is just as great, if not greater, a determiner of the kind of results obtained from this kind of research. Art educators who are interested in comparing and testing new strategies of teaching must take ample consideration of personalities involved, psychological environment, and learning theories employed.

From the findings of the present study and in regard to other discussed research results, the evidence is quite strong that involvement in depth, both quantitatively and qualitatively, must be guaranteed before a "true" measure of subjects' "spontaneity" levels can be achieved. However, a question remains as to what aspects of art learning need involvement in depth. One such area which seems to be of utmost importance is the development of aesthetic awareness.

It certainly seems reasonable that enriched aesthetic awareness presupposes enriched perceptual awareness of one's environment. The expression "meaningful art experiences" is bandied around in nearly

all discussions on art education, and it is certainly one of the most prevalent expressions seen in art education literature today. In many cases, the reader or listener is apparently supposed to know what this expression means. This writer sees meaningful art experiences as being those experiences which have the greatest potential for leading a child to his optimal level of sensitivity with his environment. These experiences lead to enriched resources which are basic to an artistic act and to creative artistic behavior. This writer would go as far as saying that student interest and enjoyment, although desirable, are not necessary criteria in the determination of meaningful art experiences. The only necessary criteria would be those based on the attributes of the artistic act and the characteristics of artistic behavior which would necessarily have to be modified for a particular group (class or grade) and in the light of their particular needs and the preparation for meeting developing needs.

Quantitatively and qualitatively, this means developing visual literacy. Visual literacy is a richer concept than visual acuity and/or awareness. Visual acuity and awareness are necessary for visual literacy and understanding. To be visually literate requires an understanding of the objective reality of one's environment and why and how one's subjective "reality" affects his perception of the environment (environment is all that exists outside of the skin and would naturally include art making). Visual literacy would include the ability to imaginatively link and embrace this "inner" and "outer"

reality. Visual literacy, in part, is determined by the extent one can gain empathy with his environment. Furthermore, the degree of visual literacy one can hope to achieve is dependent upon the extent that one can infuse a given visual experience with different thinking and empathically respond to the resulting manifestations of such infusions. Finally, visual literacy requires the development of vocabulary for the classification of visual concepts.

Investigators and writers of perception tell us that we seek stability in our visual world; however, when it comes to choosing between an object or its contextural environment as to which will take precedence in establishing visual stability, giving definiteness to the object takes precedence over the stability of the background. In the teaching of art to young or novice students, one of the difficulties that the teacher is forced to overcome is the students' "object-mindedness.". In the terminology of art this is the problem of "figure/ground" relationships. It is difficult to make these students visually aware that it is the background which gives an object its particular quality; that is, a flower in the field has a different visual quality than a flower in a vase within a still-life composition.

In order to satisfactorily express visually one's reaction to a stimulus, it follows that an individual must be aware of the aspects of the visual field which contribute to an object's unique quality. Teachers should give students analytical experiences of observing an object in multiple environments. It should not be overlooked, however,

that perceptual enrichment would include all aspects of the object also, that is, the observation and analysis of the characteristics of both its inner structure and surface features.

It is quite obvious that visual stimuli also convey qualities and these qualities are personal reactions of the viewer. It would be difficult for a teacher to get his students to learn and to accept visual qualities as he sees them; nor is this a desirable practice. After all, it is a personal response which is asked of students in art. Students need to recognize, first of all, that they have personal responses toward their environment. They need to be led into learning experiences which will give them insight into how and why personal reactions "color" perceptual experiencing.

To say a teacher should not impose his reactions to visual experience is not to say his reactions and interpretations are not important; however, several viewpoints should be offered. David Manzella, in a paper delivered at the Second Invitational Conference on Elementary Education, proposed a course which would add immensely to perceptual flexibility.[11] This course would be conducted entirely out of school and consist of the class criss-crossing the city by foot. Leaders of the group would alternately be drawn from such disciplines

---

[11]David Manzella, "The Highest Priority," The Education of Elementary School Teachers (Proceedings Second Invitational Conference on Elementary Education), R. D. Armstrong and P. A. Lane (eds.), University of Alberta, 1967.

as art, architecture, anthropology, sociology, psychology, city planning, and theology. Each leader would select and comment on aspects of the landscape based on his special strength and concerns. Although the course is designed for training teachers in art, modifications could be made for younger students. For the teacher, there are many possibilities for similar approaches to different visual stimuli.

Quantitatively and qualitatively, then, the development of visual literacy entails many different visual experiences and different ways of looking at the visual stimuli. It seems reasonable to this writer that these kinds of resources are needed before one can infuse a given visual experience with different thinking and empathically respond to the resulting manifestations of such infusions.

In theory, increased percepts lead to richer concepts. This is both a circular and spiral kind of growth, and it seems quite possible that with the enrichment of a concept there is a greater possibility for an individual to initiate or recognize new ways of perceiving. This may result and be recognized as a creative or an artistic act.

Another way of viewing the importance of developing visual literacy is related to overcoming psychological barriers to creative, artistic production. McFee attempted to formulate a learning theory (Perception-Delineation Theory) for art based upon understandings being

developed within the behavioral sciences.[12] She introduced such psychological terms into the professional nomenclature of art education as perceptual constancies, functional fixedness, response-sets, and discontinuity--all potential psychological barriers to creative artistic production. Although the author set the pattern for overcoming these obstacles, it is Arthur Efland's elaboration of the Perception-Delineation Theory in terms of learning sets which offers an answer to possible criticism of the kinds of experiences recommended by this writer. He states:

> . . . set would appear to have more consequences than might have been realized when it was first included in the development of perception-delineation. It is more than a previously acquired disposition to respond to visual information. It is also a significant aspect of the substantive content of art instruction. In other words, what the student should gain from his instruction (if he has not done so in his previous experience) is that ability to take on appropriate attending sets.
>
> Learning to do this is, in fact, learning how to learn art.[13]

Or alternatively, learning visual literacy is learning how to perceive; and becoming aware of and gaining fluency in a number of ways of perceiving (sets).

The kind of depth experiences suggested by this writer could draw criticism that the time factor would severely limit any breadth

---

[12]June K. McFee, Preparation for Art, Belmont, California: Wadsworth Publishing Co., 1961.

[13]Arthur D. Efland, "An Examination of Perception-Delineation Theory: Some Proposed Modifications," Studies in Art Education, 8:66-86, No. 2, Spring, 1967, p. 71.

of experiences. It is important to understand that the purpose of such an experience is not getting-to-know a specific object or subject but how one gets-to-know his environment. There has been much comment on visual deprivation of slum areas; however, many sensitive artists are challenging this concept. It is not that a particular environ lacks visual stimuli as much as it is the observer's lack of visual sensitivity for the stimuli. In this sense, a well-planned suburbia has less visual stimuli than the clutter of an industrial complex. It is something more than nostalgia which makes some people reluctant to see telephone lines go underground. The loss of the random patterns of line, the stateliness of the poles with their mysterious, black boxes, and the wonder of unheard messages contribute to a real deprivation of stimuli for the senses. The problem is generally not the lack of visual stimuli but a lack of visual literacy.

It should be apparent that visual literacy is not the unique domain of art; however, art is potentially the best equipped to develop visual literacy through education. Furthermore, what is being put forward is that art learning through the development of visual literacy is a prerequisite to creative artistic behavior and reflects directly on one's spontaneous behavior in art making. In order to express one's experiences in some visual art form and in particular to meet the definition of art (as defined by this author: to purposely interpret an experience by visually enhancing, clarifying, and organizing the qualities of that experience through manipulative materials), the

development of visual literacy in art resources is necessary, i.e., art media, forms, and processes, and the pictorial language of art or design.

Although these resource areas are individually identified and the development of visual literacy is discussed in terms of one's environment and now in terms of experiences in art, their interdependency must be recognized as all exist as part of one's environment. This interdependency must be recognized and promoted for the development of qualitative thought, which is basic to an artistic act and a quality art program. Art educators may claim that they are now developing visual literacy through the manipulatory experiences of doing art and this learning is transferrable to all of one's behavior. The theory that learning in one area is naturally transferred to another is highly questionable, and it is safe to say that the experiences children are now getting in an art environment are too often artificial, meaningless, and limiting.

There is a need to take a long, hard look at how transfer of learning can be optimized. Jerome Bruner goes beyond learning sets in his discussion of coding systems. First of all, Efland's statement that learning how to take on appropriate attending sets is, in fact, learning how to learn art is reminiscent of Bruner's thesis that in learning the structure of a discipline one learns how to learn.[14]

---

[14]Jerome Bruner, Process of Education, Washington, D.C.: Harvard University Press, 1960.

Bruner goes further by proposing that when one goes beyond the information given, one does so by virtue of being able to place a new given (information, experience, perception, etc.) in a more generic coding system and that one essentially "reads off" from learned contingent probabilities or learned principles of relating materials.[15] Much of what has been called transfer of learning can be considered the applying of learned coding systems to new events. Positive transfer happens when an appropriate system is applied to a new array of events, negative transfer is a case of either misapplication of a coding system to a new event or the lack of an applicable coding system. It follows from this that it is of the utmost importance in studying learning to understand systematically what it is that an organism learns. Bruner says that there are four conditions affecting the acquisition of a coding system: 1) set or attitude; 2) need state; 3) degree of mastery; 4) diversity of training.

One's set or attitude toward learning, whether a transient or an enduring thing will determine the degree to which one is equipped with coding systems that can be brought to bear on new situations. The importance of learning sets have been considered at some length and this condition need not be pursued further. The other three conditions affecting the acquisition of a coding system do need some elaboration.

---

[15]Jerome Bruner, "Going Beyond the Information Given," Cognition: The Colorado Symposium, Cambridge: Harvard Press, 1957, pp. 1-74.

The degree of mastery means more than repetitive exercises in a particular area of inquiry. Learning often cannot be translated into a generic form until there has been enough mastery of the specifics of the situation to permit the discovery of lower-order regularities which can be recombined into more generic coding systems; that is, the teacher must specify the conditions for searching out a generic coding system. In practice, this does not take place in an art program when it consists of differentiated art activities; it requires meaningfully structured depth-studies.

As to need state, there are some art educators who believe that certain kinds of information and art experiences should not be introduced until individual children profess a need for such information or experience. Assuming that a teacher can recognize a child's need, or the child can recognize his need, the need is generally immediate. Bruner points out that immediate needs produce a condition of high drive and under this condition a path to a goal is generally learned as this path to this goal, and it is not coded as an example of a more generic coding pattern, this kind of path to this kind of goal. In consequence, when a new situation arises the driven child does not have a generic coding system that permits him to go beyond it "insightfully." Individual needs are different and they arise at different stages of children's development, but we can be assured that the skills and thinking required for satisfying these needs are developmental and that these skills and thinking must be introduced and

developed to some extent before individual needs arise.

The problem which often exists in a breadth-oriented art program is that the relationships between the various activities are not noted and stressed nor is there any extensive relating beyond the task at hand. Diversity of training means much more than a series of unrelated art experiences. Unless one is exposed to some changes, genericizing does not seem to be stimulated. In order to understand the nature of something is to try to change it or to relate it to other things or experiences, for only in the face of changes in events or noting relationships of events does one begin to have the information necessary to abstract generic properties.

McFee, Efland, and Bruner have all suggested that an economy of means should be employed in learning what to experience in the events one experiences, and, in regard to Bruner, there is an emphasis on transference of learning. To sets and coding systems, can be added strategies as they, too, are means for learning. Although Beittel, Burkhart, and Bernheim identify two other possible work strategies, besides the "spontaneous" one, which seem to hold potential for art making, there is the question whether there are not others. The work strategies being investigated by Beittel, Burkhart, and Bernheim are limited to the making of visual art products, but there are other kinds of art learning such as visual information gathering from one's environment and the evaluation and criticism of the visual arts. Furthermore, the work strategies were derived from artistic operations

employed by some art students. It seems reasonable that there are learning strategies which can be devised and tested that are not now used in art learning or art making situations. In particular, there is the need to devise sets, coding systems, or strategies for developing visual literacy as a resource area. There are two other resource areas in which learning how to learn strategies can be investigated for developing consistency of artistic behavior: art media, forms, and processes and the pictorial language of art or design.

Experimentation and exploitation of art media, forms and processes comprise the second resource area to be considered. This resource area is considered to be over-emphasized and/or becomes meaningless in practice. Experimentation and exploitation should be a basic and continuous orientation of approach to new art media, forms, and processes and in pursuance of a sensitivity of their intrinsic values. This area needs structuring in order to economize the time devoted to it and to give it purpose. Too often experimentation is devoted to a few sessions of "fooling around" with art materials. This is wasted time. Exploitation of experimental results is often limited to one art project, at best, and then forgotten. This is wasted time. Exploitation is the weakest part of this orientation because experimentation becomes an end rather than a means to an end. This is not a denial of the importance of experimenting and the exploiting of these experiments for the continuing growth and development

within this resource area.

This is the process development area within and between such forms of art as drawing, easel and mural painting, printmaking, sculpting, commercial and advertizing art, etc. Process as used in this instance should be viewed as the collection of raw data with little conscious, objective, critical evaluation being made during the experience of experimenting (this is not to say that the experimenter should not be sensitive to and cognizant of how results were obtained). Examples of this raw data (processes) would be joining, bonding, bending, laminating, casting, carving, extending, filing, hammering, drilling, cutting, turning, stamping, piercing, drilling, patting, pinching, blowing, sanding, blotting, soaking, tearing, etc.

The exploitation of these experimental results requires sensitivity to the intrinsic qualities of these results and to the qualities which manifest themselves when the results are combined in terms of their relatedness or opposition. In other words, exploitation requires some degree of qualitative intelligence. This is a circular process, however, because it is assumed that exercises in qualitative thinking will increase one's capacity for qualitative thought, at least, in relation to the visual arts. It must also be recognized that the sensitizing of the sensory modules is basic to the development of qualitative thinking in the visual arts. Such sensory (perceptual) awareness requires continuous and developmental experiences

in the use of the touch/feel, sight, smell, taste, and hearing senses. These are all ways of "getting-to-know" something.

To say that experiences in experimenting are in some instances overemphasized may be misleading; what is over-emphasized is non-purposeful experimenting. It is being suggested, first of all, that it should be a continuous orientation and secondly, economy of time can be introduced by giving structure to this activity of experimenting and exploiting.

In the most simple experimental situation, there are generally three variables in art making which can be manipulated. They are the instrument, the surface, and the medium (crayolas and prepared scratchboard are two instance in which two variables are combined into one; however, it does not materially affect the usefulness of this trichotomy). Like experimental situations in the sciences, any two variables can be held constant while the other variable is manipulated; that is, a number of variations can be tried by taking one variable and using it or modifying it in many ways. However, unlike experimental situations in the sciences no variable needs to be held constant for valid results. Any single variation of one variable can be exploited for its peculiar quality, or any combination of variations within or between variables can be exploited for their related or opposing qualities, or any combination of variations within and between variables can be experimented with further and exploited when new,

emerging qualities are noted in such combinations. The possible number of experiments and exploitations approach infinity with the introduction and combination of other art forms, instruments, surfaces, and media.

The exploitation of experimental results serves two purposes: (1) it is a source of artistic inspiration by noting a particular quality or a variation of qualities in related or opposing combination of variations, and (2) it serves as a resource pool for visual recall when a particular quality or combination of qualities is being sought.

This structure can lead to the development of visual literacy but only in the very specialized area of art. The ultimate of sophistication in art is achieved when an artist can sensitively convey or express observed qualities in his general physical environment through the quality of his artistic delineation. The quality of artistic delineation is relative to the degree one has investigated the potential of his chosen art form through the manipulation of its instruments, surfaces, and media. The point is that one cannot expect to develop artistry or artistic behavior by developing one resource area at the expense of others.

Another resource area which needs greater consideration is the pictorial language of art. Design is too often considered as a separate unit for an hour or two of instruction or when there seems to be a need in a particular instance. Design is often taught haphazardly and without continuity. Art instruction, as it exists today, is the only

language which is taught with a major emphasis on the media with so little thought given to intensifying the experiences a child has as he interacts with his environment and how he can express (communicate) and order these experiences with clarity. There is a need for children to have a good basic understanding of design.

The space devoted to this resource area is not indicative of its degree of importance. Instruction in design and the development of design sensitivity are at least as important as either of the other resource areas. There are many excellent strategies for making students design conscious, and quite likely most art teachers have had worthwhile designing experiences in their education which are modifiable for teaching in lower educational levels.

Let it suffice to note, first of all, that instruction in and the continuing practice of using the design principles, organizational devices, and the visual components of art are necessary for any degree of success in clarifying, enhancing, and organizing visually the quality or essence of an experience. Secondly, it should be recognized that the ultimate goal of design experiences is to make the sense of good design "second nature." Thirdly, it should be noted of its use in structuring and understanding more "intuitive" efforts. Lastly, and certainly not least, design can be a source of discovery for new emerging qualities either through the act of manipulating the visual components of art or through the exploiting, combining, and recombining qualities discovered by experimentation. Designing

contributes to the development of visual literacy.

This portion of the discussion on the reliability of the spontaneity measure was initiated by suggestive findings in the testing of the secondary hypotheses. Depth of involvement with the object or problem was identified as characteristic of those art tasks in which sufficiently high correlations were obtained between the rank-orders of their products to warrant tentative conclusions and recommendations; that is, this variable seemed to contribute to the consistency of the subjects' "spontaneity" level between art tasks. It is hypothesized that the subjects' level of "spontaneity" must be fairly well guaranteed before other systematic variables can be adequately tested.

If it can be accepted that the visual arts reflect Man's reaction to his environment, then the scope of education through art is beyond the comprehension of any one man, and the directions which quality art programs could take offer an infinite number of possibilities. However, it has been suggested that there are at least three general areas worthy of exploiting for the development of consistency in art making behavior: 1) the development of such resource areas as visual literacy, the potentialities and limitations of various art forms, media, and processes, and the pictorial language of art; 2) the critical selection of potential art offerings from contemporary art and art education by analyzing how these offerings contribute to the development of particularly desirable abilities and to initiate

and test problem-solving experiences in relation to art materials and processes which would contribute to meaningful education through art; and 3) the investigation and testing in action learning sets, coding systems, strategies, or the like which would offer an economy of means for learning what to experience in the events one experiences and which would guarantee the optimal amount of learning transference from one art experience to other kinds of related experiences.

Although significant correlations were obtained only from the employment of the "Six Criterion Elements of Spontaneity" scale, it is being proposed by this writer that the development and exploitation of these three areas of art learning will not only contribute to the development and consistency of the spontaneity levels of subjects, but it will also contribute to the development of their creative potential as defined by the "Typology of Creativity" scale. The failure of obtaining similar results as obtained in the employment of the spontaneity measure is thought to be due in part to the lack of refinement of the "Typology of Creativity" scale; however, even with the ultimate in scales for measuring the creativeness of art products, one can hardly expect to find the same degree of consistency of "creative" behavior as one would of "spontaneous" behavior. There is just too much evidence against expecting consistent levels of creative behavior from any group of subjects. The underlying conclusion of this comparison between "creative" behavior and "spontaneous" behavior is that the two behaviors are not the same. This brings up

the question of the validity of the two criterion measures; that is, that part of validity which is concerned with the relevance of the criterion measures.

The dominant question which comes to mind concerning the "Six Criterion Elements of Spontaneity" is whether spontaneity, per se., exists only in observable behavior. Certainly an artist can be spontaneous in his thinking but need not display spontaneity in handling art materials to be identified as creative. The orginators of this criterion measure developed their criterion from observing college art education majors and their strategies in artistic problem solving.[16] A criticism of this procedure has been pointed out by Chapman:

> Again assuming that I can be set straight on these matters, there is the further problem of judging whether student responses in the research setting are influenced by their attempts at second-guessing the kind of behavior the researchers will judge to be creative. It is difficult to see how juniors majoring in art education who have done any reading in the subject area of art education or creativity (or who are in the least aware of the contemporary art scene) could not know that certain forms of behavior are currently considered more creative than others, and hence more desirable to adopt as one's own. I think it fair to say, too, that the predominate aesthetic orientation of college art departments still favors abstract expressionism and, hence, student work that appears to be "fresh and spontaneous." The reaction against the academic restraints

---

[16] Kenneth R. Beittel and R. C. Burkhart, "Strategies of Spontaneous, Divergent, and Academic Art Students," Studies in Art Education, 5:20-41, No. 1 (1963).

> of abstract expressionism is seen elsewhere, however, in "a return to the figure" and magic realism at one extreme and the development of "Pop" art at the other.[17]

Not to mention the "hard-edge" artists.

Junior high school students are more oriented toward getting-to-know their environment than in becoming spontaneously proficient in executing their art productions. The kind of spontaneity which may be prevalent in a college art studio is a rarity among early adolescent art students, and it is often an imposed condition on the part of the teacher. In this sense, spontaneity in the handling of art materials results in an artificial environment which serves as ends and does not serve as means-to-ends which are important and relevant to the junior high school students' needs.

Furthermore, the originators of the spontaneity measure used questionable validation practices in its development; they correlated art product evaluations with personality tests which were not standardized. There is the question, also, whether the authors interpreted the spontaneous strategy of their subjects from judgments of art products or from the pencil and paper test scores of the subjects.

There is less question on the inner-validity of the spontaneity measure; that is, the criterion appears thoroughly differentiated and defined so that there is little room for an individual judge to form

---

[17]Laura H. Chapman, "Some Comments on 'Spontaneous, Divergent, and Academic Art Students,'" Studies in Art Education, 6:25-29, No. 1, (1964), p. 28.

his own subjective definition and the elements of the criterion actually measure what they are defined to do. The procedure used in developing the elements of spontaneity was, first of all, the employment of a research team to name elements or parts of a representational sample of paintings which could be signs of spontaneity. The elements deducted by this process had to be logically related to the over-all conception of spontaneity, immediately understandable to the judge, and had to be applicable to any sample of two-dimensional art products. Secondly, intercorrelations were computed between elements to ascertain whether they were part of the Spontaneity Domain and additive parts of the domain. By not correlating with other domains (Aesthetic or Divergent), the elements could then be considered part of the Spontaneity Domain and non-overlapping with other domains or not contaminated with some unknown, global consideration.

There is even more doubt as to the validity of the "Typology of Creativity." The author was primarily concerned whether inter-judge agreement could be achieved in the employment of the measure in an evaluation situation and determining discrete types of creativity. As to determining discrete types of creativity, the intervalidity of the measure was not completely successful. One of the reasons for using the ranking procedure for measurement in this study was the overlap between creative types found on Eisner's drawing task.

In spite of the positively significant correlation between the two judge groups in employing the criterion measures in rank-ordering products of the same art task, there are many questions as to whether the criterion measures identify and measure the creativeness of art products. Whether the two criterion measures identify different kinds of creativeness, or one criterion measure is valid and the other invalid, or neither is valid was not satisfactorily proven by this study--nor was this a purpose of the study. What this writer is willing to admit is the validity of the spontaneity measure to evaluate what it is defined to do; that is, it measures the degree of spontaneity of execution as reflected by the surface qualities of an art product. There is a real question whether it measures the creativeness of the art product. Due credit must be given to Bernheim, the last investigator of spontaneity who has reported research results, for not alluding to or laying claim to the value of the spontaneity measure for identifying creative talent in the visual arts. As to the value of the "Typology of Creativity" as a valid measure for identifying and measuring the creativeness of art products, it depends upon one's willingness to accept the defining characteristics of the measure. The "Typology of Creativity needs to be validated and, as has been discussed previously, it needs to be made much more reliable.

This study may seem to have been an "exercise in futility;" however, it did emphasize the problems of devising criterion scales for measuring particular aspects of art products. It was pointed out

earlier in this dissertation that the first step in evaluation is to ascertain that all judges agree in their objective descriptions of what is before them. This step must be taken before any interpretation or valuing can take place. This would seem to be simple enough, but it is not the case in art. For example, consider the problem of distinguishing between two very basic elements of art: line and shape. Is a telephone pole or a window painted with one brush stroke--a line or a shape? Are Stuart Davis's delineated configurations of numerical and written symbology lines or shapes? This is one of the problems of the most simplified descriptive scales such as the scale developed by Rouse.[18] Lewis and Mussen in their evaluation of the Rouse's scale point out a number of other problems related to the employment of this scale.[19] The point is that if difficulties are encountered in just obtaining agreement among observors as to what is a line or a shape, how can agreement be expected among judges in describing relationships--let alone interpreting or valuing presented relationships?

---

[18]Mary Rouse, The Development and Validation of Descriptive Scale for Measurement of Art Products, Cooperative Research Project No. S-077, Cooperative Research Program of the Office of Education, Department of Health, Education and Welfare, 1965.

[19]Hilda P. Lewis and Paul H. Mussen, "The Development of an Instrument for Evaluating Children's Artistic Creativity," Studies in Art Education, 10:25-48, No. 3, Spring, 1969.

A judge who is expected to make a valid interpretation of the interrelationship of the visual components of an art object is faced with almost insurmountable difficulties. Does he make an interpretation of the purposes underlying the interrelatedness of an art object's visual components on the basis of his own subjective reasoning? Is this kind of interpretation valid? It seems reasonable that the judge would have to make the interpretation in relation to the maker of the art object if the interpretation is to be valid. Space is much too limited to review the research which points out the difficulties of objectively describing the complexities of any human being let along interpreting his reasons for producing a particular art work or his underlying rationale for a particular artistic arrangement. Yet, this kind of information and interpretation is needed before realistic valuing can be expected. This kind of information and interpretation is needed before "essential" criteria for evaluating art work can be isolated and refined.

One further aspect of this study needs to be reviewed and that has to do with whether there is a single art task and particular media which will elicit one's "true" level of "spontaneity" or "creativeness." A dominant characteristic of the evaluation results of this study was the degree of variance across art tasks. True, this does not give absolute proof that there is not one art task and one medium which are highly effective vehicles for bringing forth the subjects' developmental level or potential of creative artistic

ability. However, such a possibility is so remote as to make it untenable at the present time. The results of this study are a very small part of the continually increasing evidence that other variables influence the ratings of subjects' art products in any specific instance and result in a great deal of fluctuation of the ratings over a period of time. The amount and kind of art experiences, familiarity and quality of experiences with the medium, nature of the assigned task, personality factors, environmental setting, and the learning and personality dynamics in perceiving are but a few variables which could influence the art product rating at any one time or from time to time.

It has often been said that care should be taken not to push the conclusions of any research beyond what the original data warrants. The author hopes that it is clear where conclusions end and theorizing and philosophizing takes over. This writer firmly believes that those who are privileged with the intelligence to conduct research have the obligation to relate their findings to similar research results and to extend their findings as to their potential worth or possible implementation in more practical matters. Some of the connections of the writer's theorizing may appear to be, figuratively speaking, made of "hairpins" and "rusty barbwire," but their worthiness will not be known until they are tested under fire.

CHAPTER VII

IMPLICATIONS FOR ART EDUCATION

The final selection of 22 eighth grade students enrolled in a junior high school art class served as a sample of a school's population which was considered to be heterogeneous in terms of socio-economic and scholarship attributes. Furthermore, these students were at the end of their schooling in which formalized art training had been minimal and at the beginning of the period of schooling in which art instruction by qualified art teachers was available. Over a five month period the students produced art products on the basis of eight assigned art tasks which incorporated a variety of art forms, processes, and materials. The products of these art tasks served as reflectors of the artistic qualities of the students and were judged by the independent use of two measures, the "Typology of Creativity" and the "Six Elements of Spontaneity."

It was evident that the students varied widely in their performance in art; the evaluation of an individual student's performance on any one art task would have led to an erroneous prediction, in most cases, as to his performance level on other art tasks on the basis of either of the criterion measures. Although not statistically supported, a great deal of confidence would be put on the generalization that similar results would be found from studies of artistic performance of other junior high school age children.

It is concluded that goal-directed behavior in art cannot be expected or maintained and that a stable level of consistent performance cannot be achieved until art students have some confidence in their ability to handle all aspects of the artistic act. As long as the energy and attention of the students must be directed toward getting-to-know and gaining confidence in their environment, whether it is the physical or the artistic, spontaneous or creative performance levels in art cannot be estimated. This conclusion is buttressed by the evidence from this study which suggested that depth of experiencing contributed to the consistency of the subjects' "spontaneity" level across art tasks.

Furthermore, any hope investigators have in satisfactorily investigating such inductively derived variables as space dimensionality, limitations and potentialities of art media, or sensorial empathy with the art media is apparently contingent upon the subjects' consistency of artistic performance. This would probably hold true of most studies which would attempt to deductively determine variables which influence artistic performance.

As a contribution to the development of consistency of artistic performance, it is recommended that more thought and effort be given to the development of visual literacy because of its coextensive and positive relationship with the development of artistic performance. In general, this would mean that much more of the art curriculum should be devoted to: (1) depth experiences into the subjective and objective perceptual understandings of the physical and artistic environment; (2) depth experiences in relating and integrating the objective and subjective

understandings of perceiving; (3) depth of kinesthetic experiences for developing empathical sensitivity of the physical and artistic environment; (4) depth experiences in infusing a given stimulus with different and new ways of perceiving and responding, verbally and visually, to the manifestations resulting from these infusions; and (5) development of a standardized vocabulary for developing enriched perceptual and artistic concepts and for effective verbal and non-verbal communication. These kinds of experiences must be made available in both the students' everyday physical environment and that very special environment of the art classroom. Any hope of becoming sophisticated in art production necessitates the development of visual literacy in both of these environs.

Furthermore, it is recommended that the development of consistent artistic performance can be achieved by paying more attention to practices now commonly incorporated in art programming. For instance, more thought needs to be given to (1) structuring and giving purposiveness to the experimentation and exploitation experiences with art materials and (2) strategies for developing and maintaining design sensitivity.

Work strategies as defined by Beittel, Burkhart, and Bernheim are particularly attractive in conception for the development of consistency of artistic performance; however, whether the concern is with work strategies, learning sets, or coding systems, the initiation, employment, and testing of these learning systems must go beyond the limitation of being concerned only with the process of visual delineation. There is a need for devising such systems for developing resource areas which are basic for satisfactory and consistent performance in art making. Such

resource areas as visual literacy, design sensitivity, and quality experiences with art materials were discussed earlier in this dissertation; however, there are other aspects of the art learning act which need attending and one such area is the evaluation process.

Evaluation as a process and a learning experience has been largely ignored in art education instruction and research. There is a need to investigate the evaluation process during visual information handling and, in particular, the on-going process of choice making and verifying that takes place during the art making act. In relation to this kind of investigation, there is a need to devise diagnostic instruments for analyzing the on-going evaluation processes employed by children during art making. Diagnostic analysis is a prerequisite to meaningful remedial instruction. This means that the commonly accepted philosophy of "hands-off" during children's art making experiences must be modified. Finally, the evaluation process as a learning strategy, learning set, or for purposes of coding needs to be pursued through innovative and well designated research studies.

There are a multitude of art forms, media, processes, and activities which can be offered to children, and it is in keeping with and basic to the recommendations which have been set forth that attention be paid to the potential for learning that can take place from the pursuance of these art experiences; that is, the learning that takes place other than the obvious learning of a particular art process or how to make a specific art product (for example, visual literacy). Furthermore, there is a need to think through and test the learning that can be potentially achieved

through the critical selection of art experiences and possible modifications of commonly accepted approaches of introducing art forms, media, and processes. This kind of information is needed before successful coding systems can be devised and the transference of learning can be maximized (at least in the art learning environ).

The recommendation that more concern should be shown for the development of visual literacy is not limited to the purpose of conducting research. Helen Merritt has stated that we do not know what needs a child will have in science on maturity, nor can we predict what his aesthetic needs will be.[1] The development of visual literacy may be the best that can be hoped for in preparing children for that day.

This investigator cannot recommend either of the criterion measures used in this study for identifying and measuring the creativeness of art products. It follows that the criterion measures cannot be recommended for selecting creative students for gifted art classes. There are many questions as to the reliability and validity of the "Typology of Creativity"; however, the foremost question is whether a consistent level of creative performance can be maintained in art making. If it cannot, then creative students cannot be reliably identified nor can creative development be reliably measured. Although the typology seems to meet the problem of measuring the creativeness of art products head-on, the measure lacks validation in terms of relevance.

During the course of this study, the Typology Judges reported that

---

[1]Helen Merritt, Guiding Free Expression in Children's Art, New York: Holt, Rinehart and Winston, Inc., 1965.

nearly all art products judged were identified within the lowest level of creativity (Boundary Pushing) and primarily within the locus of subject (subject matter). It may well be that the art products used in this study for judging purposes were pretty uncreative, but another possibility is that the instrument itself is incapable really of differentiating within or between levels of creativity either within or across different art tasks and/or media. Although it cannot be stated with any degree of certainty whether it was the quality of the art products or the instrument which caused differentiating difficulties in judging, the writer is inclined to put the blame on the instrument. For one thing, some evidence was obtained from the training sessions of the Typology Judges that there is a real difficulty in ranking art products when they are all identified as being within the highest level of creativity (Boundary Breaking). This was the case of the products of the non-objective painting art task when the judges were called upon to determine if one art product was more "utterly new" than another. To a lesser degree, this same problem was also encountered throughout the evaluation period because nearly all art products were identified within the lowest level of creativity (Boundary Pushing). However, in the case of the non-objective art task another phenomenon was observed. The same population of subjects produced the products of the non-objective painting art task and two other art tasks which comprised part of the samples of art products used for the training of judges in the employment of the two criterion measures, but it was only in the evaluation results of the products of the non-objective painting art task that any product was judged

to have reached the highest level of creativity as defined by the "Typology of Creativity," and, then, all products of this art task were judged to be within the "Boundary Breaking" level of creativity. Briefly, in review, the process used in the non-objective painting art task militated against the development of any preconceived images that the students may have had; that is, each step of the process had to be completed before the nature of the succeeding step could be determined, and, although a few non-objective shapes were drawn on the surface at the beginning, it was quite possible these shapes would be completely different in the final product. Coupled with the relative lack of control inherent in the media used, wax and transparent inks, the final art products displayed forms (shapes) as being, by Eisner's definition of the "Boundary Breaking" level of creativity, "utterly new." The only reasonable conclusion which could be made is that the nature of the art task and media determined the "creativeness" of the products, and that the art products did not reflect the students' "true" level of creativity.

Furthermore, it must be argued that the art products used in this study were not generally uncreative in spite of the fact that most products were judged as falling within the lowest level of creativeness. In general, the art products seemed well-differentiated and reflected the heterogeneous behavioral characteristics of the subjects and the "open-endedness" of the art problems which had been presented to them.

Finally, it has been suggested that some truly creative ideas may not be attempted or consummated because of the lack of understanding of the potentialities and limitations of an art medium used by subjects in

executing their ideas. This problem was apparently encountered by the subjects in this study when they attempted to execute crayon designs with papier mache. This kind of problem does not reflect upon the capability of the "Typology of Creativity" to differentiate within or between levels of creativity; however, it does point to the need to evaluate subjects during the entire process of artistic creation.

In sum, the recommendation in regard to the "Typology of Creativity" is that this criterion measure must be refined until it is capable of differentiating finer degrees within and between levels of creativity before confidence can be put on its usefulness as a measure of the creativeness of art products.

The criterion measure "Six Criterion Elements of Spontaneity" is more reliable and valid as long as it is limited to its definition: a measure of spontaneity of surface treatment of art products. It is agreed that one of the personality characteristics of a creative person in art is spontaneity, but it does not follow that a creative artist will display his spontaneity in his artistic execution. This measure may have had some relevance in measuring creativity when Abstract Expressionism was making such an impact on art, but its relevancy has faded with this style of art. This does not mean that this criterion measure has no value. Spontaneity in handling art materials is a desirable goal in art instruction no matter what "style" a student is developing. This is especially true of children. The "Six Criterion Elements of Spontaneity" is a valuable tool for measuring the development or maintenance of spontaneity in handling art materials.

The initiation and development of such measures as the "Six Criterion Elements of Spontaneity" draw attention to the extent to which popular art movements influence what is valued and taught in art. It should be a concern of art educators to identify and evaluate these influences as to their value in art learning.

Although the difficulties in developing reliable and valid art scales seem insurmountable, effort in the development of evaluative devices for art must be continued--it is the only way that growth in art learning and the effectiveness of art programs can be reliably determined. However, much more effort must be devoted to descriptive measures. In the way of devising descriptive measures, there is a need to objectively analyze and describe process and product results from initiating experimental art learning treatments or art programming changes under various environmental conditions and with differently defined populations of subjects. Art educators would do well to approach research in art education with the same attitude as they have in developing their expertise in the manipulation of art media and processes; that is, to achieve the maximum understanding of the artistic act, it must be objectively analyzed under a variety of induced variables and conditions. It is only in the face of changes in structure and process and noting relationships within and between emerging and final phenomena that one begins to have the information necessary for describing common and discrete properties. This kind of investigation must be fairly extensive before interpretive scales can be devised which are of any value--much less having any hope of devising scales for identifying the "essential" qualities of

artistic performance.

Although the two criterion measures may hold some promise as a means of evaluating the effectiveness of one's teaching or art programming, it is suggested, at the present time, that there are more valid and reliable means of identifying children who seem to possess exceptional creative artistic potential and in predicting their chances of successfully realizing their potential in further art learning.

In general, it would seem highly desirable for any proposed program for gifted students to be able to make the best possible estimate of the potential candidates' need-to-achieve and interest levels. However, the two criteria which seem most important for identifying potentially gifted children in art are (1) visual literacy which is undoubtedly a co-requisite to artistic performance in the visual arts and (2) the ability to be fluent in gnerating ideas which is a co-requisite to creative performance.

To elaborate, and assuming that the most common reason for identifying children who seem to possess a high degree of creative artisitic potential is for the purpose of selecting children for special art instruction, it would seem highly desirable to be able to optimally estimate the children's achievement potential and interest in art learning as a means of predicting the degree of involvement and success that could be expected of possible candidates in special art instruction.

Although achievement quotients derived from the results of employing such tests as the California Achievement Tests and, as yet, the not normatively standardized Preschool Attainment Record for very young children would reflect, generally, the children's desire to achieve, it would seem

reasonable and more meaningful than relying on results obtained from verbal tests to encourage children to produce a variety of art works as a basis for evaluating their potential for successful involvement and achievement in art learning.[2,3]

In estimating interest levels of potential candidates, it would be highly desirable that they would have had a fair amount of quality experience in art; however, this is often not the case. Interest inventories, such as the Kuder Preference Record, Form A (personal), cannot be recommended because they are too imprecise for selecting gifted children for special art experiences.[4] In regard to the Kuder Preference Record, the recommended reading list of art books is dated; however, there is merit in compiling a list of selected readings about art for children to read for the purpose of developing a realistic understanding of the contemporary art scene. In either case, preference data from interest inventories or professed interest from reading art books, it would be undesirable to use this information in determining interest levels in art making. However, there are other ways of determing interest levels. For instance, in relation to the art experiences offered to children in determining their

---

[2]E. W. Tiegs and W. W. Clark, California Achievement Tests, Monterey, California: California Test Bureau (Division of McGraw-Hill Book Co.), 1957 ed., 1963 Printing.

[3]Edgar A. Doll, Preschool Attainment Record (Research Edition), Circle Pines, Minnesota: American Guidance Service, Inc., 1966.

[4]G. Fedrik Kuder, Kuder Preference Record (Form A), Chicago: Science Research Associates, Inc., 1953.

potential involvement and achievement in a special art program, a more reasonable course of action would be to interview the children upon completion of the art tasks in order to determine their degree of interest in further art training.

Continuing with the recommendation that the children should be encouraged to produce a variety of art works as a means of evaluating their potential for successful growth in special art programs, the two criteria, visual literacy and fluency, which seem most important to creative artistic performance, can be used as criterion measures for evaluating the children's in-process performance and their art works.

Quantitatively, the evaluation of the process and results of this art producing should be concerned with the fluency of ideas generated by the children in terms of art subjects and content and from discoveries made and exploited in the process of manipulating art materials. Qualitatively, the evaluation should be concerned with the kind and degree of insight the children display in the process and results of art producing; that is, the children's ability to discriminate between and relate aspects of their historical and momentary visual experiencing within both the physical and artistic environment.

Criteria for evaluating the kind and degree of insight subjects display in the process and products of their art producing could be established by making definitions of visual literacy operational or by employing existing definitions of perception which have been made operational for research purposes. For example, Marianne Frostig has not only developed tests of visual perception but also offers a program for the development

of visual perception.[5;6] The program and tests are applicable for children four to eight years of age and involve the development of five operationally defined perceptual skills: (1) perception of spatial relationships, (2) perception of position in space, (3) perceptual constancies, (4) visual motor coordination, and (5) figure-ground perception. Rather than use the Frostig development program, the interest would be in the operationally defined perceptual skills as evaluative criteria.

One point which needs to be stressed is that special art classes are initiated for particular purposes, and these purposes should determine what selection methods will be used; that is, special art classes are not formed just for the potentially gifted in art, and the selection methods suggested in the foregoing comments refer specifically to the selection of potentially gifted subjects. It should also be understood that the same criteria used for selection purposes can be used to determine the effectiveness of the selection methods and the art program during or upon completion of the special art learning experiences.

For example, a special art class may be initiated for identified underachievers. Obviously, their need-to-achieve would be low and such criteria as interest in art making, visual literacy, and fluency could be difficult

---

[5]Marianne Frostig, Marianne Frostig Developmental Program in Visual Perception, Chicago: Follett Publishing Co., 1964.

[6]Marianne Frostig and David Horne, The Frostig Program for the Development of Visual Perception, Chicago: Follett Publishing Co., 1966.

to estimate. It may be that the only indication as to how to reach these children would be from observing their outside-of-school interests. It may be for some of these children that there is an observed interest in using manipulative materials. In this case, the potential interest in manipulative materials may be the only criterion used for selecting children for a special art class for underachievers.

The program of art experiences for underachievers would be necessarily different from art experiences offered to the potentially gifted in the visual arts. This suggests another possible consideration in selection methods; that is, the nature of the art program to be initiated in a proposed special art class could enter into the determination of selection criteria. If, for instance, the art program is to be media-oriented, then those children who display an empathical sensitivity and responsiveness to art media may be the most desirable candidates for the program. On the other hand, if the program of art experiences should have an "industrial design" orientation, then children who show interest and an afinity for space design, tension and compression strengths of materials, and three-dimensional materials in general may benefit most from this program of art experiences and would be so selected.

In most cases, the suggested selection methods have an additional requirement of carefully trained judges. Art teachers experienced with the age/grade level of children being considered for a special art program would be requisite because the judges would be called upon to evaluate the quantity and quality of ideas generated and the degree of apparent insight into the art experiences; that is, the ages of children

being judged and the kind of art media, processes, and tasks offered to the children would determine expected outcomes, and informed, experienced art teachers would be more sensitive to these expectancies which, in turn, would serve as a basis for making more valid, value judgments.

Furthermore, it is recommended on the basis of this investigator's experiences during the course of this study that the judgments should not be limited to the art products. It is important that judges begin their observations upon the initiation of art tasks and to pay particular attention to the children's original ideas and intentions. The purpose would be not only to determine the fluency and quality of the children's ideas but also to determine the reasons for any succeeding modifications of their ideas and intentions during the art experience. Obviously, the children's reasons for changes of intent or ideas could be quite varied and not very predictable; however, there are two kinds of process change which would be particularly informative and valuable for judges to take into account in their evaluations.

First, procedural changes may be due to the lack of information, skills, and resources which children at a particular age-level could not be expected to possess. Obviously, a shift to less innovative ideas due to these reasons should not lower the judge's estimate of these children, and, in such cases, initial ideas and intentions should be the source of value judgments. Second, procedural changes which reflect the children's flexibility in overcoming in-process problems, awareness and exploitation of emerging relationships, and sensitivity and enhancement of discovered expressive qualities in manipulating a particular medium

would be a valuable indicator of these children achieving the maximum benefits from art instruction in special art programs. The point which needs to be stressed is that the art product is the result, and not a record, of what took place in the process of art making.

By carefully selecting and training judges in gathering and evaluating the kinds of information which has been recommended they look for, one could use their trained judgments as criterion measures of the children's artistic interest and fluency in art making.

Furthermore, the foregoing recommendation (that in the employment of art judges there should be particular attention paid to their selection and training) can be generalized to the use of art judges in research. From discussions held with the judges in this study and the experience with teachers as judges, it is recommended that more concern should be shown in the personality, art training, and art production interests of judges used in research studies. And, to reiterate, more research should be devoted to evaluation as a process of information-handling during the act of teaching, learning, and judging with particular concern for the process of decision-making and verifying that children employ in the process of art making. Furthermore, art educators who are concerned with the preparation of art teachers would do well to teach prospective art teachers as much as possible about the process of evaluation, not only in the judgment of art products but also in evaluating the art motivations and art processes included in art programming.

SUGGESTIONS FOR FURTHER STUDY

Evidence, and lack of evidence, from this study suggests research in the following areas:

1. Longitudinal studies for comparing the two criterion measures used in this study with other kinds of populations. Of particular interest would be a longitudinal evaluation of art products of subjects who are in a special art class or who have had much more experience in and aptitude for art than was the case of the population in this study.

2. Longitudinal studies need to be made for comparing or investigating art scales as to their reliability and validity in operational situations in the schools.

3. Continued attempts must be made to develop reliable and valid descriptive scales for inventorying art products. It would be worthwhile to attempt further differentiation and specification between and within levels of creativity in the visual arts.

4. Further research needs to be conducted in judge characteristics, particularly the influence of training and interest in art on the judges' perceptual ordering and emphases of the defining components of a criterion scale.

5. Scales need to be developed for measuring complexity of art products and, in particular, in relation to visual literacy.

6. Such inductively derived variables as space dimensionality, limitations of the media, and sensorial empathy with the media were not satisfactorily determined, one way or another, as to their influence on artistic performance. Further study needs to be conducted with a population which displays more consistent art performance.

7. There are a number of assumptions and reasoned conclusions in the proposed development of visual literacy which are worthy of testing. Of particular value would be trying out treatments which may contribute to visual literacy.

8. Investigations should continue into strategies, learning sets, and coding systems which would economize and maximize art learning.

9. Studies need to be conducted into the ways children behave in manipulating various art media and to determine the value of this behavior for achieving desirable art learning goals.

10. Research needs to be conducted into the process of evaluation for the purpose of improving the teaching and learning acts in art. Of interest, would be the possibility of identifying learning sets or coding systems in evaluating which would contribute to more effective and productive evaluation processes.

BIBLIOGRAPHY

Alford, Mary Lee. "Teacher Judgments as Related to Certain Predictors of Artistic Creativity in Senior High School Students." Unpublished doctoral dissertation, North Texas State University, 1964.

Barkan, M. and Hausman, J. "Two Pilot Studies for the Purpose of Clarifying Hypotheses for Research into Creative Behavior." Research in Art Education, Seventh Yearbook. Washington, D. C.: National Art Education Association, 1956.

Barron, F. and Welsh, B. S. "Artistic Preference as a Factor in Personality Style: Its Measurement by a Figure Preference Test." Journal of Psychology. XXXIII (1952), 199-203.

Beittel, Kenneth R. "In Reply to Laura Chapman's Critique." Studies in Art Education. VI, No. 1 (1964), 30.

Beittel, Kenneth R. and Burkhart, R. C. "Strategies of Spontaneous, Divergent, and Academic Art Students." Studies in Art Education. V, No. 1 (1963), 20-41.

Beittel, Kenneth R. and Mattel, Edward L. "The Effect of a 'Depth' vs. a 'Breadth' Method of Art Instruction at the Ninth Grade Level." Readings in Art Education. Edited by Elliot W. Eisner and David W. Ecker. Waltham, Mass.: Blaisdell Publishing Co., 1966.

Bernheim, Gloria. "The Dimensionality of Differential Criteria in the Art Product." Studies in Art Education. VI, No. 1 (1964), 31-48.

Bruner, Jerome. "Going Beyond the Information Given." Cognition: The Colorado Symposium. Cambridge: Harvard Press, 1957.

Bruner, Jerome. Process of Education. Washington, D. C.: Harvard University Press, 1960.

Burkhart, R. C. Spontaneous and Deliberate Ways of Learning in Art. Scranton: International Textbook Co., 1962.

Chapman, Laura H. "Some Comments on 'Spontaneous, Divergent, and Academic Art Student'." Studies in Art Education. VI, No. 1 (1964), 25-29.

Doll, Edgar A. Preschool Attainment Record (Research Edition). Circle Pines, Minesota: American Guidance Service, Inc., 1966.

Dubin, Robert. "Deviant Behavior and Social Structure: Continuities in Social Theory." American Sociological Review. XXIV, No. 2 (1959), 147-164.

Edmonston, Paul. "Overview: Product and Evaluation." Art Education Bulletin. Kutztown: Eastern Arts Association. XVIII, No. 4 (1961), 51-59.

Efland, Arthur D. "An Examination of Perception-Delineation Theory: Some Proposed Modifications." Studies in Art Education. VIII, No. 2 (1967), 66-86.

Eisner, Elliot. "A Typology of Creativity in the Visual Arts." Studies in Art Education. IV, No. 2 (1962), 11-22.

Eysenck, H. J. Dimensions of Personality. London: Routledge, Kegan Paul, 1947.

Frankston, Leon. "Some Explorations of the Effect of Creative Visual Art Experiences upon the Poetry Writing Quality of Eighth Grade Students." Studies in Art Education. V, No. 1 (1963), 42-59.

Fritzky, Ferdinand. Aesthetic Preference for Abstract Designs as a Function of their Perceived Complexity. Princeton N. J.: Educational Testing Service, RB-63-27, 1963.

Frostig, Marianne. Marianne Frostig Developmental Program in Visual Perception. Chicago: Follett Publishing Co., 1964.

Frostig, Marianne and Horne, David. The Frostig Program for the Development of Visual Perception. Chicago: Follett Publishing Co., 1966.

Goodenough, Florence L. Measurement of Intelligence by Drawings. New York: Harcourt, Brace, and World, 1926.

Goodenough, Florence L. and Harris, Dale B. "Studies in the Psychology of Children's Drawings: II, 1928-1949." Psychology Bulletin. XLVII (1950), 369-433.

Harris, Dale B. Children's Drawings as a Measure of Intellectual Maturity. New York: Harcourt, Brace, and World, 1963.

Holland, Jean. "A Developmental Scale for Grades 6, 7, 8." Unpublished research. Toronto School System: Toronto, Canada, 1958.

Klein, Stephen P. A Description of Points of View in Esthetic Judgments in Terms of Similarity Dimensions. Princeton, N. J.: Educational Testing Service, RB-67-53, 1967.

Kuder, Fedrick G. Kuder Preference Record (Form A). Chicago: Science Research Associates Inc., 1953.

Lansing, Kenneth. "The Effect of Class Size and Room Size upon the Creative Drawings of Fifth Grade Children." Research in Art Education, Ninth Yearbook. Washington, D. C.: National Art Education Association, 1959.

Lantz, Beatrice. Easel Age Scale. Los Angeles: California Test Bureau, 1955.

Linderman, Earl. "The Relation of Art Judgment to Judge Personality." Studies in Art Education. III, No. 2 (1962), 46-51.

Lowenfeld, Viktor. Creative and Mental Growth, 3rd Edition. New York: The MacMillan Company, 1957.

MacGregor, Ronald N. "Imposed Controls in Subject Matter and Art Media Choice." Alberta Journal of Educational Research. XX, No. 2 (1967), 103-110.

Machover, Karen. Personality Projection in the Drawings of the Human Figure. Springfield: O. C. Thomas, 1949.

Manzella, David. "The Highest Priority." Proceedings of the Second Invitational Conference on Elementary Education, Department of Elementary Education, University of Alberta, 1967.

Maslow, A. H. "A General Theory of Personality Research." An abstract from a speech (via radio) at the Interdisciplinary Symposium. Michigan State University, 1959.

Maslow, A. H. Motivation and Personality. New York: Harper and Brothers, 1954.

McFee, June K. "Creative Problem Solving Abilities of Academically Superior Adolescents." Unpublished research supported by a Ford Foundation grant, Stanford University (mimeographed).

McFee, June K. Preparation for Art. Belmont, California: Wadsworth Publishing Co., 1961.

Merritt, Helen. Guiding Free Expression in Children's Art. New York: Holt, Rinehart and Winston, Inc., 1965.

Merton, Robert K. Social Theory and Social Structure. Glencoe, Illinois: Free Press, 1957.

Owen, Cramer. "An Investigation of Creative Potential at the Junior High Level." Studies in Art Education. III, No. 2 (1962), 16-33.

Roe, Ann. The Psychology of Occupations. New York: John Wiley and Sons, 1956.

Rouse, Mary. The Development and Validation of Descriptive Scale for Measurement of Art Products. Office of Education, Department of Health, Education and Welfare, Cooperative Research Project No. S-077, 1965.

Siegal, Sidney. Nonparametric Statistics for the Behavioral Sciences. New York: McGraw Hill Book Company, Inc., 1956.

Silverman, Ronald H. "Comparing the Effects of Two- and Three-Dimensional Art Activity upon Spatial Visualization, Aesthetic Judgment, and Art Interest." Unpublished doctoral dissertation, Stanford University, 1962.

Tiegs, E. W. and Clark, W. W. California Achievement Tests. Monterey, California: California Test Bureau, 1957 Edition, 1963 Printing.

Torrance, E. Paul. Minnesota Tests of Creative Thinking. Kent State University: Bureau of Educational Research, 1964.

Torrance, E. Paul. Torrance Tests of Creative Thinking. New Jersey: Personnel Press, Inc., 1966.

Witkinn, Herman. "Perception of the Upright." Scientific American. February, 1959.

Wolff, W. "Projective Methods for Personality Analysis of Expressive Behavior in Preschool Children." Character and Personality. X, No. 4 (1949), 309-330.

APPENDIX A

SIX CRITERION ELEMENTS OF SPONTANEITY

The following elements of spontaneity are not listed in order of merit; each is equally important as a guide for evaluation. All elements of spontaneity will not be characteristic of all art tasks, nor will all be found in all art products. It is the sum degree of spontaneity which determines the rank order of any art product.

(1) Objects; Moving to Still. Focus on objects as shapes and make a judgment as to whether they look static or moving with the latter being of the highest value.

(2) Movement within Shapes. Focus on shapes within the object and make a judgment as to the amount of movement which is apparent. The greater movement the higher the value.

(3) Detail; Loose to Meticulous. The amount of detail is not considered. High ratings are given to loose detail.

(4) Media Overlap. Look only at the use of the media and determine if there is much (high) or little (low) overlapping of media.

(5) Border of Shapes: Ragged to Smooth. The greater degree of raggedness of borders the higher the value of spontaneity.

APPENDIX A - Continued

(6) Texture; Artist and Media Interacting to Mechanically Imposed. The mechanically created textural effects are given the lowest value.

APPENDIX B

TYPOLOGY OF CREATIVITY

Three types of creativity and two loci constitute the classes of the typology. The three classes determine the level of creativeness with boundary breaking being the highest, inventing next, and boundary pushing being of the lowest order. The extent to which characteristics within a certain class are displayed in an art product also serves to determine the rank order of an art product in its relation to other art products displaying the same characteristics.

(1) Boundary Breaking

(a) Subject. The production of new subject matter through the creation of the completely new or through reversing the premises upon which old subject matter was developed.

(b) Form. The formulation of utterly new forms.

(2) Inventing

(a) Subject. The production of new subject matter through the combination of known subject matter.

(b) Form. The production of new forms through the combination of known forms.

APPENDIX B - Continued

(3) Boundary Pushing

(a) Subject. The extension of ordinary subject matter through novel combinations of such subject matter or through novel elaboration.

(b) Form. The extension of common forms by the novel combination of such forms or by their novel elaboration.

APPENDIX C

DESCRIPTION OF ANCILLARY TESTS USED IN THE DEVELOPMENT OF THE SPONTANEITY MEASURE

1. VM, or Visual Magila: An experimental exercise for visual preference patterns. This test consists of 32 items, each of which contains five pictures from which two are to be chosen on the basis of preference. Pictures have the flavor of nonrepresentational process samples in black and white media. Each item has samples representing a process-control dimension; a masculine-feminine dimension (square, heavy, bold, dark, etc., as opposed to round, light, graceful, etc.); and a unique organizational dimension. These dimensions are so combined that there is in each item the following types of samples: (1) process-feminine, (2) process-masculine, (3) control-feminine, (4) control-masculine, and (5) unique organization.

2. Process Blackness Application (IBM). This is an actual IBM test score derived from inserting students' drawings made with IBM pencils on the Nitschke Drawing Behavior Samples Test into an IBM test scoring machine for blackness count.

3. Process Involvement Recall from Product Stimulus. There is and A and B form of this checklist on each of which 50 adjectives descriptive of process feelings and thoughts are presented for checking through recall on the student's part with his art object before him.

4. Word Selection Preference Patterns Test (SDT). This is an 42-item test in which five words are presented to the subject on each item. As in the VM, the respondent is asked to choose only two. The five words represent choices originally labeled: (1) abstract orientation, (2) process orientation, (3) theory orientation, (4) impulsivity orientation, and (5) nonconventionality orientation.

5. Process Independence and Self-reliance (PIT). This is a test which presents to the student on each item five statements representing decisions or attitudes concerning independent work and the working process, and asks the student to rank these.

APPENDIX C - Continued

6. ISSD, or Test of Ideational and Social Self-determination. This is a test which has gone through three revisions and has proven useful in the description of creative action. There are four scales: (1) complexity, (2) theoretical interest, (3) aesthetic interest, and (4) social self-determination.

7. BBCI-X3, or Experimental Creativity Inventory. This is a 130-item, 14 "content" scale inventory. These are named: (1) originality, (2) general sensitivity, (3) flexibility, (4) interest, (5) independence, (6) action, (7) abstraction from perception, (8) inclusion of the opposite, (9) confidence (risk), (10) humor, (11) tension, (12) fluency, (13) supportive environment, (14) skill, and (15) creative orientation.

8. FAT, or Flexibility-Abstract Orientation Test. This is a test of 80 items, 40 of which are of the self-descriptive statement type and 40 of which are work pairs. The statements reflect an action orientation typified by spontaneous flexibility, whereas the word pairs reflect a preference for abstract over concrete words.

9. WET, or Word Equation Test. This is a power test asking the student to find a common denominator or principle which will order three words (e.g., floor, chair, ceiling) and give an explanation for doing so. The student is asked to give as many common denominators within a given time period as possible.

10. OQT, or Object Question or Divergent Question Test. This is a timed test in which the respondent is to ask "open" questions about common objects, such as "ice." The questions are to arouse interest and curiosity, and lead to a variety of responses which are neither factual nor capable of a "yes-no" answer.

11. RAT, or Remote Associates Test. In this test three words are presented in each item and the respondent is asked to discover a word which "mediates" a common association among them. An example is: blue, cottage, rat; which words can be mediated by the remote associate, "cheese."

12. Art Evaluative Power. This is a performance in evaluating in which the student is asked to do two things: (1) describe as many different criteria as possible that will rank eight pictorial samples and (2) find the sample of the eight most like a drawing of his own. The sample have known ranks according to spontaneous, divergent, and formal criteria.

TYPIST

Miss Minnie Rosychuk