A VALIDITY STUDY OF THE WALLACH-KOGAN CREATIVITY TEST:
THE PREDICTION OF SIX CONCURRENT CRITERIA IN VISUAL ART

DISSERTATION

Presented in Partial Fulfillment of the Requirements for
the Degree Doctor of Philosophy in the Graduate
School of The Ohio State University

By

Fred Harold Wallbrown, B.S. Ed., M. Ed.

* * * * *

The Ohio State University
1971

Approved by

[Signature]
Adviser
Department of Psychology
© 1972
FRED HAROLD WALLBROWN

ALL RIGHTS RESERVED
ACKNOWLEDGEMENTS

My greatest indebtedness is to the members of the reading committee whose interest, inspiration, and consideration made completion of the study a stimulating intellectual experience rather than an academic hurdle. In specific, I wish to express my appreciation to Dr. Charles B. Huelsman, Jr.; Dr. Robert J. Wherry, Sr.; and Dr. Ann Engin whose collective efforts contributed so much to execution of the dissertation. Their contribution can be repaid only through efforts on my part to equal their excellence as teachers, researchers, and persons.

I am deeply grateful to Dr. Huelsman, my advisor and committee chairman, not only for guidance in preparation of the dissertation itself but also for a long and profitable period of association dating back to my first two courses at Ohio State in 1965. In a very real sense Dr. Huelsman's contributions to my education and professional development are too numerous to mention. However, I am especially grateful for his constructive criticisms and suggestions which were instrumental in reshaping the study during the planning stage. Dr. Huelsman's stringent editing not only
resulted in substantial improvement in the dissertation but also provided an excellent lesson in technical writing.

First and foremost, however, I am most appreciative of Dr. Huelsman's commitment to students, effectiveness as a teacher, and implacable integrity. Dr. Huelsman always managed to find time to provide assistance despite the pressing demands of a heavy teaching load, extensive program development, and numerous advisors. This often amounted to meeting on weekends and after regular work hours. Needless to say, Dr. Huelsman was always available to provide assistance on short notice.

I am particularly indebted to Dr. Wherry whose hierarchial theory of cognition provided the conceptual framework for the study. Dr. Wherry took many hours from his busy schedule to discuss design, data analysis, and interpretation relevant to the study. Working with Dr. Wherry throughout the various phases of the study provided the kind of learning experience which can be obtained only through contact with a distinguished methodologist.

Yet, Dr. Wherry's contribution to my education extends far beyond assistance with the present study. I am deeply obligated to Dr. Wherry for his effectiveness as a teacher and commitment to students. Despite the responsibilities of a full teaching load and the department chairmanship, no student seeking assistance ever found his way to
Dr. Wherry's office blocked except to wait his turn while another student received help.

At the beginning of correlational analysis, my first course with Dr. Wherry, statistical analysis, was totally shrouded in mystery for me. The extent to which this mystery has been dispelled is directly attributable to Dr. Wherry's patience, encouragement, and effectiveness as a teacher.

Numerous other individuals also contributed to the study. Professor Glen Rothman, Department of Art, rendered an invaluable service both in serving as a content consultant and nominating judges to rate the creative products. Without the assistance so graciously provided by Professor Rothman the study could not have been implemented.

Dr. R. F. Sletto, Department of Sociology, made important contributions to the study. It was Dr. Sletto who first suggested that biographical and situational variables should be included among the predictors. The idea of using biographical variables to generate hypotheses regarding antecedents of creative performance also originated with Dr. Sletto. The perspective of a distinguished methodologist from a sister discipline was extremely valuable in suggesting leads for future creativity research.

Dr. Michael Wallach of Duke University also contributed substantially by granting permission to reproduce
copies of the W-K creativity test for use in the study. In addition, Dr. Wallach took time from a busy schedule for three extended telephone conversations concerning proper interpretation and scoring of the test.

Dr. Damon Asbury and Mr. Richard Amorose, longtime friends and colleagues, provided many hours of assistance in data collection and offered valuable suggestions regarding conduct of the study. It was, for example, Dr. Asbury who first suggested the present interpretation of the verbal precision factor.

I am also indebted to my employer, The Columbus Ohio Public Schools, for permitting flexibility in my work schedule so that data collection could be completed.

Cooperation from the parents, teachers, and children of St. John's Elementary School made conduct of the study possible. Examiners working in the school were provided with excellent facilities and remarkable flexibility in scheduling. On several occasions the teachers went so far as to voluntarily give up their lounge to provide private testing facilities. Needless to say, graduate students seeking permission to conduct research are not frequently received in such a gracious manner.

Special thanks are due for the assistance provided by Sr. Elaine, Sr. Helena, Sr. Celia, and Mrs. Lusan Ridgeway. Sr. Elaine not only arranged for conduct of the study at St. John's but also provided invaluable assistance.
in collecting biographical data not readily available from the school records. Without the assistance of Sr. Helena and Sr. Elaine collection of accurate biographical data could not have been achieved.

Sr. Celia and Mrs. Ridgeway contributed to the study in many ways. They graciously allowed almost complete freedom in scheduling children from their classes for over two months. In addition, they handled collection of creative products for children in their respective classes which involved no small expenditure of time and energy. It is quite obvious that the study could not have been conducted without the assistance of these teachers.

I am also indebted to Mrs. Lillie Parrott for her patience and perseverance in typing the various iterations of the dissertation.

Finally, heartfelt thanks are highly appropriate to my wife, Barbara, and to our children who shared intimately in all of the successes and frustrations of my graduate career.
VITA

June 4, 1937

1959

1959-1960

1960-1961

1961

1961-1962

1962-1964

1964-1965

1965-1967

1967-1968

1968-1970

Born: Spencer, West Virginia

B.S. Ed: Ohio University, Athens, Ohio

Graduate Assistant: College of Education, Ohio University, Athens, Ohio

Teacher: South-Western Local School District, Harrison, Ohio

M.Ed.: Ohio University, Athens, Ohio

Teacher-Counselor: Southern Local School District, Hemlock, Ohio

Teacher-Counselor: Yellow Springs Exempted Schools, Yellow Springs, Ohio

Elementary Counselor: South-Western City Schools, Grove City, Ohio

Research Assistant: The Ohio State University Evaluation Center, Columbus, Ohio

Intern School Psychologist: Department of Child Study and Student Counseling, Columbus Public Schools, Columbus, Ohio

School Psychologist on Special Assignment (Evaluation Specialist): Department of Evaluation and Research, Columbus, Public Schools, Columbus, Ohio
1970-1972 School Psychologist:
Department of Child Study
and Student Counseling,
Columbus Public Schools,
Columbus, Ohio

FIELDS OF STUDY

Major Field: School Psychology

Psycho-Educational Assessment Professor Charles B.
Huelsman, Jr.

Learning Disabilities Professor Charles B.
Huelsman, Jr.

Research Design and Statistics Professor R. J. Wnerry, Sr.

Methodological Sociology Professor R. F. Sletts
TABLE OF CONTENTS

ACKNOWLEDGEMENTS......................................................... ii
VITA.................................................................................. vii
LIST OF TABLES................................................................. xi
LIST OF APPENDIXES....................................................... xiii

Chapter

I. INTRODUCTION................................................................. 1
   Statement of the Problem
   Importance of the Study
   Definition of Terms

II. LITERATURE REVIEW.................................................... 7
   The Research and Theory of J. P. Guilford
   Research Based Upon the Work of Wallach
   and Kogan
   Research and Theory Concerning Criterion
   Development

III. RATIONALE FOR THE PRESENT STUDY.......................... 68
   Selection of a Creativity Instrument
   Selection of the Creativity Criteria
   Description of Predictor Variables
   Description of Criterion Variables
   Considerations in the Design of the Study

IV. METHOD AND PROCEDURE............................................. 83
   General Hypothesis
   Sample
   Data Collection Procedure
TABLE OF CONTENTS (continued)

V. RESULTS AND DISCUSSION............................... 101

Preliminary Data

Question 1: To what extent will the
discriminant validity from
general intelligence and
high subtest intercorrelations
reported by W-K and others be
generalized to pupils in the
present sample with a different
social, ethical, and economic
background?

Question 2: Will the four judges achieve
a significant degree of concordance in their ratings of the
artistic products submitted to
them for judgement?

Question 3: What is the factorial composition
of the predictor and
criterion variables employed
in the study?

Question 4: Among the variables examined;
five creativity subtests,
eleven WISC subtests, and nine
control variables; what is the
most effective combination for
predicting concurrent creative
performance in the areas of visual
art selected for study?

VI. CONCLUSIONS............................................ 191

APPENDIXES..................................................... 197

REFERENCES.................................................... 217
# LIST OF TABLES

<table>
<thead>
<tr>
<th>Table</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.</td>
<td>Notation Used for Predictor Variables</td>
<td>95</td>
</tr>
<tr>
<td>2.</td>
<td>Notation Used for Criterion Variables</td>
<td>100</td>
</tr>
<tr>
<td>3.</td>
<td>Descriptive Statistics for the Five Creativity Subtests</td>
<td>104</td>
</tr>
<tr>
<td>4.</td>
<td>Intercorrelations Among the Five Creativity Subtests</td>
<td>109</td>
</tr>
<tr>
<td>5.</td>
<td>Intercorrelations Between the Creativity and Intelligence Subtests</td>
<td>111</td>
</tr>
<tr>
<td>6.</td>
<td>Intercorrelations Among the Eleven WISC Subtests</td>
<td>113</td>
</tr>
<tr>
<td>7.</td>
<td>Interjudge Agreement for Each of the Six Creativity Criteria</td>
<td>116</td>
</tr>
<tr>
<td>8.</td>
<td>Hierarchial Factor Matrix</td>
<td>124</td>
</tr>
<tr>
<td>8.1</td>
<td>Factor g - General Intelligence</td>
<td>126</td>
</tr>
<tr>
<td>8.2</td>
<td>Factor c - General Creativity</td>
<td>132</td>
</tr>
<tr>
<td>8.3</td>
<td>Factor I - Visual Creativity</td>
<td>134</td>
</tr>
<tr>
<td>8.4</td>
<td>Factor II - Undefined</td>
<td>136</td>
</tr>
<tr>
<td>8.5</td>
<td>Factor 1 - Verbal Creativity</td>
<td>138</td>
</tr>
<tr>
<td>8.6</td>
<td>Factor 2 - Verbal Precision</td>
<td>140</td>
</tr>
<tr>
<td>8.7</td>
<td>Factor 3 - Freedom From Distractibility</td>
<td>144</td>
</tr>
<tr>
<td>8.8</td>
<td>Factor 4 - Media Specific Creativity</td>
<td>147</td>
</tr>
<tr>
<td>8.9</td>
<td>Factor 5 - Coding</td>
<td>149</td>
</tr>
<tr>
<td>8.10</td>
<td>Factor 6 - Intact Home</td>
<td>152</td>
</tr>
<tr>
<td>8.11</td>
<td>Factor 7 - Education of Father</td>
<td>154</td>
</tr>
<tr>
<td>8.12</td>
<td>Factor 8 - Perceptual Organization</td>
<td>156</td>
</tr>
</tbody>
</table>
9. Stepwise Regression Procedure for C₃  169
10. Regression Equations for C₃  170
11. Stepwise Regression Procedure for C₄  178
12. Regression Equations for C₄  179
# LIST OF APPENDIXES

<table>
<thead>
<tr>
<th>Appendix</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>Distribution of Intelligence Quotients for the Sample</td>
<td>198</td>
</tr>
<tr>
<td>B</td>
<td>Distribution of Birth Order Variable for the Sample</td>
<td>199</td>
</tr>
<tr>
<td>C</td>
<td>Distribution of Chronological Age for the Sample</td>
<td>200</td>
</tr>
<tr>
<td>D</td>
<td>Rating Procedure for Art Products</td>
<td>201</td>
</tr>
<tr>
<td>E</td>
<td>Comprehensive Data Matrix</td>
<td>204</td>
</tr>
<tr>
<td>F</td>
<td>Distribution of z-Scores for the Five Creativity Subtests</td>
<td>205</td>
</tr>
<tr>
<td>G</td>
<td>Predictor-Criterion Intercorrelation Matrix</td>
<td>206</td>
</tr>
<tr>
<td>H</td>
<td>Regression Equation for C3 Obtained from Stepwise Regression Procedure Based on Conservative Decision Rule</td>
<td>207</td>
</tr>
<tr>
<td>I</td>
<td>Regression Equation for C4 Obtained from Stepwise Regression Procedure Based on Conservative Decision Rule</td>
<td>208</td>
</tr>
<tr>
<td>J</td>
<td>Full Stepwise Regression Procedure for C1</td>
<td>209</td>
</tr>
<tr>
<td>K</td>
<td>Regression Equation for C1 Obtained from Full Stepwise Regression Procedure</td>
<td>210</td>
</tr>
<tr>
<td>L</td>
<td>Full Stepwise Regression Procedure for C2</td>
<td>211</td>
</tr>
<tr>
<td>M</td>
<td>Regression Equation for C2 Obtained from Full Stepwise Regression Procedure</td>
<td>212</td>
</tr>
</tbody>
</table>
N  Full Stepwise Regression Procedure for C5  213
O  Regression Equation for C5 Obtained from Full Stepwise Regression Procedure  214
P  Full Stepwise Regression Procedure for C6  215
Q  Regression Equation for C6 Obtained from Full Stepwise Regression Procedure  216
CHAPTER I

INTRODUCTION

The decade of the 1960's has been marked by a gradual increase in the quantity of research concerning creative children and youth. Among the topics of primary concern to researchers have been the identification of creative children, the psychological and sociological correlates of creativity, education programs for the development of creativity, the differences between creative and gifted children, and the effect of teacher behavior on creativity in children. Despite the increase in the quantity of research, very little is known about the creative child. In fact, researchers have not been able to agree on what a creative child is like.

Several factors appear to account for the confusion with regard to creativity research. First, there is no widespread agreement as to an appropriate definition of creativity. Second, most tests purported to measure creativity are devoid of both sound psychometric properties and a firm theoretical base. Third, few researchers have attempted to validate the test they develop as creativity measures against a criterion of creative behavior.

An operational orientation provides a convenient strategy for avoiding the morass of verbal speculation which
has so frequently characterized creativity research. In specific, an anchor point can be obtained by formulating an operational definition of creativity in accordance with the research strategies discussed by Underwood (1957). The writings of Campbell and Fiske (1959) as well as Underwood (1957) clearly suggest that one should select a reliable set of creativity operations which show discriminant validity from general intelligence and define a stable dimension of cognition.

It should be clearly understood that the variety of operationalism advocated by Underwood (1957) is in no way antagonistic toward psychological theory. Instead, operationalism constitutes the first step in the development of psychological theory. The researcher with an operational orientation is free to develop his own theoretical constructs, relate his findings to existing theories, or continue to work at an operational level. In any case, operationalism has the advantage of focusing attention on the distinction between theoretical constructs and the operations used to make inferences regarding these constructs.
Statement of the Problem

It was the purpose of this study: (1) to provide a replication of selected aspects of the Wallach and Kogan (1965) study employing a sample with diverse background characteristics, (2) to examine the extent to which individuals trained in art agree in their ratings of selected artistic products of children, (3) to determine the factorial composition of the predictor and criterion variables employed in the study, (4) to determine the most effective combination of variables (from the 25 employed in the study) for the prediction of creative performance in selected areas of visual art.

Importance of the Study

The most important aspect of the present study is that it represents a criterion-based approach to the validation of a set of operations designed to measure creative potential. This type of research is basic to further progress in the area of creativity research. Researchers such as Martinson and Seagoe (1967), Brodén and Sprecher (1964), Taylor (1964), and Taylor and Ellison (1964) have called attention to the need for criterion development in creativity research. Unfortunately, the advice of such authorities has received little attention from creativity researchers.
A second important feature of the present study consists of the incorporation of a wide range of predictor variables. Available research and theory suggested the following three classes of promising predictor variables: creativity test, intelligence tests, and biographical information. Variables from all three classes were included in the study in order that their relative contribution to the prediction of the creativity criteria could be objectively determined.

Furthermore, the present study represents a logical extension of the research initiated by Wallach and Kogan (1965) and continued by researchers such as W. Ward (1966, 1968), Pankove (1967), Cropley (1968), Cronbach (1968), J. Ward (1967, 1968), Fee (1968), Wallach and Wing (1969), and Elkin, Deblinger and Adler (1970). That is, this study represents a logical next step in investigating the validity of the Wallach-Kogan Creativity Test and indirectly the constructs upon which the test was developed.

The research noted above indicates that Wallach and Kogan (1965) successfully developed a set of creativity operations which possess both convergent and discriminant validity and a reasonable degree of reliability. When such a state of knowledge has been attained with regard to a psychological trait, the next logical step consists in determining the range of validity for that trait. One of the most crucial types of validity for a creativity test is that of pre-
dictive validity. Predictive validity consists of determining the extent to which the test correlates with various creativity criteria across a broad range of samples.

Definition of Terms

The lack of agreement regarding a satisfactory definition of creativity is readily evident from even a cursory examination of the literature. For example, Yamamoto (1965), Frierson (1969), and Chiselin (1963) have called attention to the consequences of this disagreement as to an appropriate definition of creativity. The problem of definition goes even deeper, however. Methodologists such as Underwood (1957), Lawson (1965), and Campbell and Fiske (1959) have clearly shown that agreement as to a literary definition of a concept does not necessarily lead to the development of congruent or even similar instruments for the measurement of that concept. A good example of this occurrence in the creativity literature is Mednick's (1962) definition of creativity. Mednick used his definition of creativity to develop the Remote Associates Test. Wallach and Kogan (1965) accepted Mednick's definition of creativity but developed an entirely different set of creativity operations from the same definition.

Consequently, the problems of formulating a creativity definition satisfactory to all, or even many creativity
Consequently, the problems of formulating a creativity definition satisfactory to all, or even many creativity researchers appears to defy solution at the present time. As noted above, there is also the associated problem of moving from concept to operation. The theoretical formulations of Underwood (1957) and Campbell and Fiske (1959) provide an alternative to the problems associated with formulating a satisfactory definition of creativity. That is, one can adopt an enlightened operationalism which escapes the logical flaws inherent in strict operationalism. The flaws of strict operationalism are well documented by Kaplan (1963, pp. 39-42). However, the concepts advanced by Underwood (1957) and further developed by Campbell and Fisk (1959) establish procedures whereby operations can be tested for equivalence thus avoiding the logical flaws of strict operationalism.

Within the context of the above discussion, creativity will be defined as that which is measured by the Wallach-Kogan Creativity Test. That is, the operations performed in securing scores for this test will define the concept of creativity as it is employed in the present study.
CHAPTER II
LITERATURE REVIEW

The present chapter is devoted to an examination of the creativity literature most relevant to the present study. Two major considerations governed the selection of the literature for discussion. First, literature essential to the rationale for the study was selected. Second, literature was selected in such a manner as to illustrate the relationship of the present study to the broader area of creativity research. Obviously, the present review is far from exhaustive since numerous volumes have been devoted to creativity research.

Emphasis is given to three major aspects of the creativity literature. The first section is devoted to an examination of the creativity research based upon divergent production tests related to Guilford's structure-of-intellect model. The next section is devoted to an examination of research and theory related to Wallach and Kogan's (W-K) creativity operations. Finally, the third section is devoted to an examination of the problems associated with criterion development.
Research Based Upon J. P. Guilford's
Structure-of-intellect Model

The Research and Theory of
J. P. Guilford

The theoretical formulations of J. P. Guilford (1950, 1956, 1957, 1959, 1967) have exercised a profound influence on the evolution of creativity research. The structure-of-intellect model was first announced by Guilford in 1956 and has undergone frequent elaboration since that time. The empirical basis for the model was derived from Guilford's experience in the factor analysis of intelligence tests. Guilford's methodological preferences led him to reject the hierarchial models from the british tradition in favor of a morphological model.

In fact, Guilford (1967) developed a firm set of methodological preferences in the course of pursuing factor analytic studies designed to complete his taxonomy of intellectual abilities. First, Guilford has expressed a preference for including simple rather than complex tests in his factor analyses so that individual differences in total scores for a test reflect one unique ability and minimize variance in all other abilities. Second, Guilford has expressed mistrust of analytic rotation methods utilizing computers preferring to use subjectively guided, graphic rotations.

Third, Guilford regards intellectual factors as logically orthogonal and has persistently refused to employ
techniques involving oblique solutions or even an hierarchi-
al solution which maintains orthogonality, e.g., Wherry (1959).
Thus, Guilford's theoretical preferences preclude considera-
tion of higher-order factors necessary for the development
of hierarchial models of intelligence. Guilford's rationale
for exclusive concern with orthogonal primary factors is illu-
strated by the following quotation:

There is little point in seeking to learn
about higher-order factors when such fac-
tors rest upon information about inter-
correlations of first-order factors, whose
values are uncertain (1967, p. 471).

Fourth, Guilford has expressed a preference for
speed as opposed to power in developing tests for completion
of his structure-of-intellect model. Guilford employs speed
as an experimental control to prevent the examinee from making
translations and substitutions instead of relying on the pri-
mary ability required by the test.

The structure-of-intellect model contains three
parameters: operations, products, and content. The present
form of the theory makes provision for five kinds of opera-
tions--cognition, memory, divergent production, convergent
production, and evaluation. Four kinds of content--figural,
symbolic, semantic, and behavioral--comprise the second dimen-
sion of the model. The third dimension of the model consists
of products which may occur in any of the following six forms:
units, classes, relations, systems, transformations, and im-
plications. By way of explanation, it should be noted that
Guilford defines a product as a form in which information may occur. Thus, the structure-of-intellect model makes provision for 120 cells or combinations of operations, products, and content. The ordering of categories along each of the three dimensions of the model was determined on a logical rather than empirical basis.

Guilford’s work may be viewed as an extension of the task which Thurstone set for himself, i.e., "... to delineate the various domains of individual differences in what are usually called 'cognitive' abilities, and from the results to make inferences concerning the psychological processes involved in cognitive tasks" (Carroll, 1968, p. 249). Consequently, the research efforts of Guilford and his associates have been concentrated more on the development of tests to complete the 120 cells of the model than on the conduct of studies to demonstrate the validity of these tests for predicting criteria in the real world (Guilford, 1964). Considerable progress has been made in developing tests to define the cells of the model. Guilford (1967) reported evidence for the existence of 81 unique intellectual abilities which occupy 77 cells of the model. The number of abilities identified exceeds the number of cells filled because of the involvement of sensory modes.

For the purposes of the present discussion, the major concerns are with Guilford’s construct of divergent production because of its impact upon creativity research.
The publication of Guilford's (1956) structure-of-intellect model heightened interest in creativity research by calling attention to hypothetical types of abilities not included in conventional intelligence tests. At the same time, his tests of divergent production provided researchers with instrumentation for the measurement of divergent production.

At an operational level, Guilford's tests of divergent production are of the completion form in which the examinee is required to produce a number of plausible responses which satisfy a set of specifications. Responses are scored according to: 1) number of responses, 2) uniqueness of responses, or 3) some quality criterion. Tests of divergent production are distinguished from traditional ability tests in that there is no one "right" answer for the task. Instead, the examinee earns a high score on the basis of the criteria mentioned above. Guilford characterized tests requiring the examinee to produce one correct answer as tests of convergent ability.

At the empirical level, the divergent production tests developed by Guilford and his associates have not held up well in validity studies involving correlations with creativity criteria. In fact, the few empirical studies employing children as subjects suggest that tests of divergent thinking do not correlate significantly with either achievement or creativity criteria. For example, Cicirelli (1965) found that creativity, when added to the regression equation for intelligence and achievement, did not significantly raise
the level of prediction. In other words, a traditional intelligence test was found to predict classroom achievement relatively well, but addition to the creativity battery did not increase the predictability of achievement.

Martinson and Seagoe (1967) compared the effectiveness of the Stanford-Binet (Terman & Merrill, 1960) and five of the Guilford divergent production tasks in discriminating between pupils whose creative products had been rated high and low by a group of content specialists. The Binet successfully discriminated between the high and low groups on five of the eight criterion measures. That is, pupils with high Binet scores were rated significantly higher by judges on their creative productions than pupils with low Binet scores. However, the Guilford test failed to discriminate between pupils rated high and low on their creative products. Alternatively, when high and low groups were formed on the basis of scores on the Guilford tests, no significant differences were found on any of the criterion variables.

At the adult level, studies investigating the criterion validity of the Guilford tests of divergent thinking have yielded equivocal results. However, when one considers the strength of relationship found in studies supposedly showing positive results, it seems relatively safe to assert that the validity of the divergent production tests remains to be established.

On the positive side, Drevdal (1956) correlated
scores on a series of divergent production tests with judges' ratings of creativity on the part of college students and obtained a modest correlation of 0.33 between originality scores and criterion ratings. Barron (1955) found that three divergent production tests correlated in the 0.30's with staff assessment of originality on the part of a sample of Air Force officers. Elliott (1964) also obtained results supporting the criterion validity of the Guilford divergent production tests. Two groups of public relations officials—one high and one low in creativity—were identified by their supervisors' ratings. Elliott found that five of the eight tests employed were able to discriminate between these two groups.

Zaccaria, et al. (1956) obtained data to support the criterion validity of the Guilford tests in a unique manner. This group combined signs of creative performance derived from biographical data into a creativity criterion and found significant correlations between 11 of 15 divergent production tests and the criterion index.

In addition to the studies discussed above, there are several studies employing teacher and peer ratings of creativity as criteria which provide support for the criterion validity of divergent production tests. Lauritzen (1963) found that teachers' ratings of originality correlated 0.48 with originality as measured by divergent production tests. However, teacher ratings did not correlate
significantly with fluency as indicated by the divergent production tasks employed. Most investigations employing teacher ratings of creativity as criteria have yielded more equivocal results than the Lauritzen study.

On the negative side, studies by Beittel (1964) and Skager, *et al.* (1967) failed to obtain significant correlations between divergent production scores and ratings of performance in art. MacKinnon (1961) found that the divergent production tests employed in his study of architects did not correlate significantly with criterion ratings of creativity as indicated by experts in architecture or an index derived from numbers of publications. In a similar study, Gough (1961) found that tests of divergent production correlated negatively with creativity ratings for research scientists. Negative results were also obtained by Jacobsen and Asher (1963) who found inconsequential relationships between tests of divergent production and a work sample criterion with a sample of college students.

The sample of studies cited above illustrates the conflict of evidence concerning the validity of Guilford's tests of divergent production. In view of the negative findings and weak relationships found in many studies reporting positive results, conservative decision rules would indicate suspending judgement at the present time. However, the burden of proof rests with Guilford and his associates. Much more research is necessary to determine the validity parameters of
Guilford's divergent production tests. The viewpoint stated above is congruent with the stance adopted by MacKinnon (1961, p. 32) who concluded that, "In view of such negative findings the use of Guilford's battery of creativity tests for the identification of creative persons would be, to say the least, questionable." Dellas and Gaier (1969) reached essentially the same conclusion after an extensive review of the literature involving validity studies of these tests. They stated that: "The results, thus far, have been contradictory and far from conclusive" (p. 56).

Research Involving Modification of Guilford's Work

Much creativity research has been conducted with slight modifications of Guilford's divergent production tests. This tendency was especially pronounced with regard to creativity research involving children as subjects. During the early 1960's there was a proliferation of divergent production tests devised for use with children. Guilford's original tasks were often modified in some small way to suit a given researcher's notions about the nature of creativity.

Among the first studies to employ such a modified divergent production test with school age children was that of Getzels and Jackson (1962). Using a sample of 533 pupils from grades 5-12, these authors identified a highly creative group and a highly intelligent group and proceeded to compare and contrast the psychological characteristics and traits
of these two groups. The selection procedure employed by Getzels and Jackson eliminated pupils who were both highly creative and highly intelligent since their aim was to differentiate between these two groups. The creativity battery employed by Getzels and Jackson consisted of five tests derived, directly or indirectly, from Guilford's work. The tests comprising the battery were purported to provide a measure of divergent thinking. Scores for the five tests were summed to obtain a composite score.

The results of the Getzels and Jackson (1962) study indicated that the creative group was distinguished from the highly intelligent group by a wider range of interest, greater emotional stability, and a better sense of humor. Both groups were about equal (did not differ significantly) in standardized test performance in basic school subjects even though the high intelligence group had a mean IQ score 23 points higher than the creative group. Teachers were found to prefer having the high intelligence group in their classes rather than the creative pupils. In general, the Getzels and Jackson study reported socially desirable differences in favor of highly creative pupils over highly intelligent pupils.

Another early study was conducted by Flescher (1963) who studied the relative contributions of creativity and intelligence to classroom achievement. The Flescher study represented one of the earlier attempts to establish
validity for a creativity battery. The Getzels and Jackson study was used as a basis for planning. However, several changes were made. The Getzels and Jackson tests were modified, and the number increased to seven. A sample of 110 sixth grade pupils of both sexes was used.

Flescher found that the intelligence test scores correlated significantly with achievement in all areas. In contrast, creativity tests scores were not related to achievement in any area. Measures of general anxiety and test anxiety were not significantly related to either general intelligence or creativity. The intercorrelations among Flescher's creativity subtests were low. The average correlation among the subtests of this creativity battery was 0.11.

Studies exploring the relationship between divergent production tests and intelligence are illustrated by the following: Cline, Richards and Needham (1963); Cline, Richards, and Abe (1962); Clark, Veldman, and Thorpe (1965); Cicirelli (1965); and Schmandel, Merrifield, and Bonsall (1965). In these studies the approach of the Getzels and Jackson (1962) study was generally followed in that children scoring high on creativity tests were contrasted with children having high IQ scores. In cases where criterion validity was a concern, these studies indicated that creativity tests similar to those originally developed by Getzels and Jackson were neither effective in predicting school
achievement singularly nor in increasing the predictive efficiency of a battery of tests designed to predict school achievement.

Studies such as those of Martinson and Seagoe (1967) and Cicirelli (1965) supported the earlier findings of Flescher (1963) by indicating that the Guilford-derived tests of divergent thinking added nothing to predictive relationship already established between traditional intelligence tests and performance criteria. In addition, both reanalyses and methodological critiques of the Getzels and Jackson study have thoroughly discredited their findings. Authorities such as Marsh (1964), Thorndike (1966), Burt (1962), and W-K (1965) showed that the tests of divergent thinking correlated as highly with measures of general intelligence as they did among themselves.

The divergent production tests of Torrance have enjoyed considerable popularity. Torrance (1966) appears to have used Guilford's concept of divergent production as a means of scoring a series of tasks which Torrance himself selected as "...models of the creative process, each involving different kinds of thinking and each contributing something unique to the battery under development" (p. 9). Torrance employed four of Guilford's divergent production factors (fluency, flexibility, originality, and elaboration) as scoring categories for his battery. Beyond this, Torrance has not explained the relationship of his work to Guilford's
structure-of-intellect model. Rather, Torrance proposed his own definition of creativity which, by implication, seems to have served as the basis for selecting the tasks to be included in his creativity battery. A recent version of Torrance's definition is as follows:

A process of beinging sensitive to problems, deficiencies, gaps in knowledge, missing elements, disharmonies, and so on: identifying the difficult; searching for solutions, making guesses, or formulating hypotheses about the deficiencies: testing and retesting these hypotheses and possibly modifying and retesting them; and finally communicating the results (1966, p. 6).

According to Torrance, this definition provides the basis for "...defining operationally the kinds of abilities, mental functioning, and personality characteristics that facilitate or inhibit the (creative) process" (p. 7). Two considerations were, according to Torrance, influential in his formulation of the above definition. These were an analysis of previous creativity definitions "...and the requirements of a definition for keeping a program of research focused on factors affecting creative growth in context..." (ibid). Furthermore, Torrance indicates that his definition is in harmony with the historical meaning of creativity and is equally applicable across a wide variety of human endeavors. A high degree of creativity as defined by Torrance's operations is, it would appear, a necessary although not sufficient condition for creative behavior in an individual's day-to-day acti-
The Torrance Tests of Creative Thinking (Torrance, 1966) represent the author's most recent attempt to operationalize his definition of creativity. The Torrance battery consists of five verbal tests and three figural tests. The five verbal tests are Ask-and-Guess, Products Improvement, Unusual Uses, Unusual Questions, and Just Suppose. The three figural tests are Picture Construction, Incomplete Figures, and Parallel Lines. The following four criteria are employed in scoring the tests of the Torrance battery: 1) fluency, 2) flexibility, 3) originality, and 4) elaboration. Fluency consists of the number of appropriate responses. Flexibility consists of the number of spontaneous shifts from one meaning category to another. Originality consists of the statistical frequency of a given response, i.e., the less frequent the response the more original. Elaboration consists of the number of different ideas or details provided.

Examination of the manual revealed that several types of validity are claimed for this test. First, content validity is claimed for the battery on the basis that it includes a wide range of tasks that previous creativity research has shown to be involved in creative behavior. The following quotation illustrates Torrance's position with regard to content validity:

To insure content validity, a consistent and deliberate effort has been made to base the test stimuli, the test tasks,
instructions, and scoring procedures on
the best theory and research now avail-

In specific, Torrance goes on to state that his creativity
tasks were selected on the basis of some combination of the
following: analysis of the lives of persons acknowledged as
creative, research concerning the personalities of creative
persons, analysis of performances generally acknowledged to
be creative, and theory and research regarding human cognition.

In his discussion of construct validity, Torrance
clearly rejects the possibility that creativity might consti-
tute a unitary cognitive dimension by stating that the
battery was constructed so that low sub-score intercorrela-
tions would be obtained. This position is congruent with
that of Guilford whose theory was the major impetus for the
construction of the Torrance battery. Given this conceptual
stance, most of the studies cited as providing construct valid-
ity for this battery consist of either comparisons of the per-
sonality characteristics of high and low scorers on the Tor-
rance tests or simple correlations between creativity test
scores and other measures. In addition, studies showing im-
proved performance on the Torrance tests after creativity
training were also cited as providing construct validity.

Research concerning concurrent validity was also
discussed at length. In this realm, most data consisted of
correlations between the Torrance tests and one of the follow-
ing variables: classroom achievement, peer ratings of crea-
tivity, and teacher ratings of creativity.

The most unfortunate characteristics of the validity research related to the Torrance tests has been a lack of concern with criterion development. A related characteristic of this validity research has been the frequent employment of simple correlation between a composite creativity index and some other variable. A much more appropriate approach would consist of determining the regression of various creativity criteria on the Torrance subtests. Use of a composite creativity score is entirely inappropriate when Torrance himself denies the existence of a unitary cognitive dimension of creativity.

In a review of the Torrance tests, Wallach (1968) provided a summary of the validity studies included in the test manual as well as other relevant research now available. In this review, Wallach cited evidence showing substantial correlations between the Torrance test and conventional intelligence tests. From this point Wallach proceeded to show that the correlations between the Torrance test and the criteria of teachers and peer ratings could be accounted for on the basis of variance shared between the Torrance tests and general intelligence.

Wallach also showed that the Torrance battery adds little power to tests of general ability in the prediction of academic achievement. The small increment in predictive power sometimes found, Wallach surmised, is probably due to the fact
that the Torrance battery provides a measure of some aspect of general intelligence not included in the conventional intelligence tests employed. Wallach summarized his review of the validity studies as follows: "...we have been able to find little evidence in support of an interpretation of the Torrance tests that would construe them as 'creative thinking' rather than simply as 'thinking'" (p. 277).

In addition, Wallach showed that Torrance's definition of creativity does not differ substantially from general intelligence as traditionally defined except that he "...adds explicit reference to a cognitive characteristic that tends to be slighted in customary definitions: the ability to engage in divergent search processes directed toward potentially useful guesses about solution" (p. 273). Wallach found little in the test battery to indicate that this divergent search process received operationalization.

An Evaluation of Research Based upon Guilford's Work

Data summarized in the previous sections clearly indicates that the validity of all classes of Guilford-derived divergent production tests remains to be established. An examination of the reasons for this state of affairs has important implications for future creativity research. Authorities such as Taylor and Holland (1964), Brogden and Sprecher (1964), and Deles and Gaier (1970) have frequently called attention to a disconcerting state of affairs which
figures prominently in the lack of validity data concerning divergent production tests. That is, researchers employing divergent production tests have showed a consistent disregard for the problem of criterion development. These researchers have made few serious attempts to develop criteria for use in evaluating creative products. The reasons for this disregard of criterion development differs somewhat from one research to another.

Three major approaches to the criterion problem can be ascertained from an examination of the literature regarding the validity of divergent production tests. The first approach is exemplified by the work of Torrance (e.g. 1966) who relies on correlations between scores on his creativity battery and convenience criteria. Among the convenience criteria reported by Torrance are teacher and peer ratings of creativity, achievement test scores, and class marks. The writings of Torrance do not make clear the rationale for his use of convenience criteria rather than examining actual creative products.

The work of Getzels and Jackson (1962) illustrates the third approach. Here the criterion problem is simply ignored. These authors simply assembled a battery of tests which they felt should measure creativity. Instead of attempting to collect validity data for their test, Getzels and Jackson skipped the step without attempting to relate scores on their test to performance on actual product-based creati-
vity criteria.

In the absence of adequate data from criterion-based studies, one must rely on other less desirable means of evaluating the various divergent production tests. Wallach (1970) proposed the dimensionality issue as the basis for such an evaluation. That is, Wallach (1970) examined the extent to which divergent production tests measure a cohesive dimension with discriminant validity from general intelligence. Analysis of available research data clearly indicated that existent divergent production tests fail to define such a unitary cognitive dimension independent of general intelligence. This state of affairs has been documented by a large number of researchers during the last decade. For example, Burt (1962), Vernon (1964), Thorndike (1963), Marsh (1964), and Wallach and Kogan (1965) have noted the lack of dimensionality at one time or another.

Writing in a broader perspective, McNemar (1964) severely attacked the naive use of unvalidated primary mental ability batteries such as those developed by Guilford. Research previously summarized clearly indicates that the intercorrelations among the subtests of existent divergent production tests are so low that postulation of a unitary cognitive dimension cannot be justified. Furthermore, the small amount of common variance shared by divergent production tests can be most parsimoniously explained by general intelligence.
This lack of convergent and discriminant validity constitutes a serious constraint to the practical use of divergent production tests. Thorndike (1963, p. 423) pointed out that the low intercorrelations among divergent production subtests means that the particular persons selected as creative will be largely dependent upon the particular types of divergent production tasks included in the selection battery.

Another probable weakness in the various divergent tests stems from their administration in an evaluative atmosphere with rather strict time limits (W-K, 1965). This position is supported by data from two distinct sources. First, the biographical data analyzed by Chiselin (1952) indicates that creative responses are most likely to require a considerable period of incubation and finally emerge through a process of combinatory or associational play. Such an atmosphere does not occur in a conventional testing situation. Second, a study by Dentler and Mackler (1964) showed that examinees perform better on a test of unusual uses under relaxed conditions than when their performance was being evaluated. Until more definitive research data become available to indicate otherwise, one is left with the nagging suspicion that an evaluative atmosphere with time limits may well function to suppress the emergence of creative responses.

The above discussion of testing conditions is closely related to a conclusion reached by Dellas and Gaier
(1970) in their recent discussion of the creativity literature. These authors expressed the opinion that a systematic study of examiner influence and testing conditions should be incorporated in future creativity research. This point could be extended to include systematic study of the time variable which Guilford uses as a method of experimental control in test development.

Another serious deficit in the research involving divergent production tests is the lack of careful replication. Replication should play a very important role in an area which is rife with contradiction and conflict in research findings. The following quotation from Dellas and Gaier (1970) is an excellent commentary on the confusion caused by the lack of careful replication in creativity research:

Paradoxically, the paucity of replication studies simultaneously exists with the plethora of literature concerning one particular investigation that is merely re-written with essentially the same information (p. 69).

The major reason for this dearth of replicative studies may well be associated with the lack of esteem with which such studies are regarded by many psychologists. In any case, more enlightened guidance of graduate students planning creativity research could ameliorate this situation in a rather short period of time.

The entire set of control variables indentified by Brogden and Sprecher (1964) and further discussed by research-
ers such as Taylor and Holland (1964), Taylor (1964), and Taylor and Ellison (1964) has been almost entirely neglected in the numerous studies involving divergent production tests. Unless the investigator exercises either experimental or statistical control over biographical, time, and opportunity variables, his research is open to serious confounding. Such errors in design cannot be defended on the basis of either inaccessibility of data or excessive time and effort because such data are routinely available in school files. Furthermore, systematic study of these data can be readily achieved through multiple regression analysis.

Finally, there has been a lack of longitudinal research employing divergent production tests. Numerous authorities such as Dellas and Gaier (1970), Taylor and Holland (1964), and Wallach (1970) have pointed out that many issues in creativity research can be resolved only by longitudinal research. Such research is not only expensive in terms of time and money but also in terms of the total life span of a given researcher. However, it would seem that researchers of such national stature as Guilford or Torrance should have access to the funds necessary to initiate longitudinal research.

Research Based Upon the Work of Wallach and Kogan

The Theoretical Basis of the
Wallach-Kogan Study

Wallach (1970) thoroughly explored the antecedents of the W-K (1965) study. Detailed exploration of this account
is beyond the scope of the present discussion. For purposes of the present study it is not inappropriate to state that the theoretical basis of the W-K study was provided by the work of Mednick (1962). In the course of his theoretical exposition, Mednick (p. 221) stated his associative hypothesis of creativity in the form of the following definition:

...we may proceed to define the creative thinking process as the forming of associative elements into new combinations which either meet specific requirements or are in some way useful.

The second aspect of Mednick's exposition consisted of an elaboration of potential individual differences among people with regard to the creative thinking process. Considerable attention was also devoted to Mednick's operationalization of his theory. However, discussion of the procedure is not directly relevant to the present discussion.

According to Mednick's theory, the organization of an individual's response hierarchy account for individual differences in creativity. The slope of the curve associated with an individual's response hierarchy is a critical variable. Mednick discusses two extreme types of response hierarchies. The first type of hierarchy is characterized by a curve with a steep response gradient. Individuals with such a hierarchy typically provide associations at a rapid rate but their reserve of associations is quickly exhausted. In addition, the responses provided by such an individual tend to be commonplace and stereotyped. Few unique or unusual responses are given.
The second type of hierarchy is characterized by a curve with a shallow response gradient. Individuals with such a hierarchy give their responses more slowly but continue giving responses for a much longer period of time. The early responses by these individuals also tend to be stereotyped and commonplace. As this type of individual continues to respond his stereotyped responses are exhausted and more unique responses appear.

The major difference between the two types of gradients is that the strength of the stereotyped responses is not so great in the case of an individual with a shallow gradient. Consequently, the more remote kinds of associations are expended. In the case of the steep gradient, we find the associative strength concentrated in a few stereotyped associations instead of being more evenly distributed across a larger number of associations.

Mednick does acknowledge that creative behavior is possible on the part of an individual with a steep response gradient. Some individuals may have a steep response hierarchy which is deviant from that of the general population. The dominant associations of such an individual will also be so strong as to prohibit emergence of more unique associations. However, the dominant responses for such a person will be unique insofar as the remainder of the population is concerned. Creative individuals with this type of associative hierarchy tend to be one-shot producers. If they are successful in producing
more than one creative product, later products will usually bear close resemblance to the first one. In contrast, an individual with a relatively flatter hierarchy is much more likely to be a multiple producer.

For individuals with steep response gradients as well as those with shallow response gradients, stereotyped associations usually have the greatest associative strength so they are emitted first. However, initial response emission is more rapid in the case of the individual with the steep response gradient because the associative strength of the stereotyped responses is greater than for the individual with the shallow gradient. When the stereotyped responses are expended by the individual with the steep gradient he is unable to provide more responses because his associative repertoire is exhausted.

Thus, it can be seen that Mednick hypothesized a creative individual should produce not only a greater number of associations but also more unique associations in response to a structured task.

According to Mednick's definition, the essence of creative thinking is the formation of associative elements into new combinations. The new combinations must meet specific criteria such as those specified in a laboratory situation or in the task demands of the real world. In other words, the new combinations must meet the criterion of utility. Thus, Mednick proposed the criteria of originality and utility for
determining creative responses. Mednick acknowledged the potential difficulty in applying these two criteria to behavior in the real world. The criteria of utility is often most difficult to attain because of the problems associated with attaining consensus as to what is useful. In a laboratory situation the experimenter can, however, arbitrarily establish and explain his criteria for usefulness to the subject. In the case of originality Mednick proposed a statistical definition phased in terms of probability of response.

In the discussion of his associative hypotheses, Mednick indicated three ways in which a creative solution may be achieved. The following quotation provides the context for understanding the nature of a creative solution:

Generally, any condition or state of the organism which will tend to bring the requisite associative elements into ideational contiguity will increase the probability and speed of a creative solution (p. 221).

The first way of achieving a creative solution is that of serendipity. That is, the necessary elements for a new combination may be evoked contiguously by a complex of stimuli from the environment. Examples of serendipity in creative thinking are the invention of the X-ray and the discovery of penicillin.

The second way of achieving a creative solution consists of similarity. "The requisite associative element may be evoked in contiguity as a result of the similarity of
the associative elements or the similarity of the stimuli eliciting these associative elements" (p. 222). Mednick surmised that similarity may be most important in areas of the arts such as painting, sculpture, and poetry where creative effort is less dependent on the direct manipulation of abstract symbols.

The third method of achieving a creative solution is that of mediation. Mednick described mediation as follows: "The requisite associative elements may be evoked in contiguity through the mediation of common elements" (p. 222). This type of creative solution is most applicable in areas of endeavor where the manipulation of abstract symbols is crucial. Examples of such areas would be mathematics, chemistry, and physics.

Thus far, Mednick's associative theory of creativity has been discussed in a normotheic sense. As a basis for deducing individual differences in creative thinking, Mednick made the following statement:

Any ability or tendency which serves to bring otherwise mutually remote ideas into contiguity will facilitate a creative solution; any ability or tendency which serves to keep remote ideas from contiguous evocation will inhibit the creative solution (p. 222).

Mednick also provided a discussion of possible personality or cognitive styles which might influence creative thinking. When a problem solution demands concrete representation of a problem an individual with a perceptual approach
would be at a disadvantage. Here Mednick seemed to be describing qualitative differences in associative hierarchies which cut across the nomothetic aspects of his theory. He goes on to discuss possible differences between visualizers and verbalizers. A verbalizer is one who approaches a problem by the use of words while the visualizer is one who conceptualizes a problem through use of sensory representations of relevance to concrete aspects of the problem.

The Wallach-Kogan Study

Using the theoretical formulations of Mednick as their point of departure, W-K developed five creativity subtests for use in their study. According to W-K's interpretation of Mednick's formulations, when an individual is provided with a general task orientation, the total number of associations and the number of unique associations should reflect an individual's creative potential. In other words, an index of creativity can be obtained by summing standard scores for these two variables.

To operationalize their interpretation of Mednick's theory, W-K borrowed freely from the tests developed by other researchers. The work of Guilford and his associates proved to be a fruitful source of ideas. However, the creativity battery finally developed by these authors possessed little in common with the creativity tests used by other researchers. Despite obvious similarities in item content, the scoring method and conditions of administration developed by
W-K represented a unique approach to the measurement of creativity.

W-K deduced the conditions for administration of their creativity battery from Mednick's theory. They reasoned that freedom from time pressure and a gamelike atmosphere were necessary conditions for the emergence of responses with low associative strength. Such responses are to be found at the low end of the shallow response gradient described by Mednick. Empirical support for this position was cited in the work of Boursefield and Sedgewick (1944); Christensen, Guilford, and Wilson (1957); and Dentler and Mackler (1964).

Another concern discussed by W-K was the failure of creativity researchers to demonstrate the existence of a unitary dimension independent of general intelligence. In the course of this discussion they drew heavily on the work of Campbell and Fisk (1959). They first established the task of determining discriminant validity from general intelligence. That is, it was necessary to establish that the creativity subtests measured a variable independent of general intelligence. Secondly, W-K established the requirement that the subtests of their creativity battery should be intercorrelated to the extent that the postulation of a unitary dimension of creativity would be justified.

The final creativity battery developed by W-K consisted of the following five subtests: Instances, Alternative Uses, Similarities, Pattern Meanings, and Line Meanings.
Each subtest was scored for both number of associations and uniqueness of associations, making ten separate scores in all. The ten raw scores from the battery were separately converted to standard scores and summed to yield a single creativity index.

W-K selected a sample of 151 fifth grade pupils (70 boys and 81 girls) which comprised the total enrollment (six classes) at this grade level in a suburban New England school system. The six classes were housed in two elementary schools. According to W-K (p. 27), the family background of the pupils was predominantly middle class, white, and Protestant. Ninety-four percent of the fathers of pupils in the sample were employed in upper-level blue collar occupations such as electrician or carpenter. No Negroes were included in the sample.

W-K chose a homogeneous social stratum for their sample to avoid contamination of their findings by differences in socio-economic status. These researchers seemed to be making the assumption that their sample was representative of middle class children of the United States. While W-K admitted the desirability of including lower socio-economic status pupils in their sample, they noted that doubling the size of the sample would "...multiply by a factor of two the several years of work our investigation required" (p. 26). In other words, excessive preoccupation with sampling concerns (in terms of available resources) would have prohibited conduct of the study.
Thus, W-K found it necessary to test the Mednick creativity theory by employing a homogeneous sample of middle class pupils and to leave the tasks of replication and generalization for future research.

Analysis of the data obtained from this sample strongly supported the research hypotheses of W-K. Split-half reliability estimates for the ten subtests scores were substantial. The average split-half reliability coefficient for the ten subtests was 0.70 with only two of the coefficients falling below 0.80. The reliability estimates for the Instances subtest were much lower than those for the remainder of the battery. That is, the average reliability for the ten subtests of the battery was considerably depressed by the Instances subtest. With regard to this subtest the reliability coefficients for number of responses was 0.75 while the reliability coefficient for uniqueness of responses was only 0.51.

An item analysis for the ten scores provided by the five subtests yielded substantial item-sum correlations. All 78 of the items correlated higher than 0.40 with the total score and 71 of the 78 item-sum correlations were 0.60 or higher. These findings indicate that all 78 items in the battery make a substantial contribution to the total score.

Furthermore, the ten subtest scores showed substantial intercorrelations. It was found that 43 of the 45 subtest intercorrelations attained significance at beyond the
0.1 level. The findings justify the summation of the standard scores for the ten subtests to obtain a creativity index. In addition, the ten subtests scores showed negligible correlations with an index of general intelligence.

The obtained dimensionality and divergence from general intelligence justified the postulation of a creativity dimension independent of general intelligence. In the words of W-K (p. 56), "Under the conditions of freedom from evaluation and absence of time pressure...the ability to generate associative verbal responses comes from a source different from intelligence as traditionally conceived."

After demonstrating that their creativity battery possessed sound psychometric properties, W-K proceeded to independently classify the pupils in their sample according to scores on the intelligence and creativity index were classified as "high creativity" while those scoring below the median for their sex were classified as "low creativity." A similar classification procedure was employed in the matter of intelligence. Pupils of a given sex who scored above the sample median on the intelligence index were classified as "high intelligence" and those scoring below the median intelligence index for their own sex were classified as "low intelligence."

A joint distribution was formed by dichotomizing pupil performance on the creativity and intelligence measures. Thus, the following four groups were constituted: high creativity-low intelligence, high creativity-high intelligence, low creativity-high intelligence, low creativity-low intelligence.
ty high intelligence, and low creativity-low intelligence.

A series of 2 x 2 analyses of variance were performed for purposes of comparing the four groups of pupils mentioned previously on several dependent variables. The broad categories of dependent variables employed were behavior in the school environment, categorizing and conceptualizing, sensitivity to physiognomic properties, and anxiety and defensiveness. Separate analyses were performed for the scores of boys and girls on all dependent variables.

In the course of their data analyses involving the four classes of variables noted above, W-K presented preliminary findings which not only support the validity of their creativity test but also suggest interesting possibilities for educational practice. These findings are not reviewed. Several considerations led to this decision. First, the findings essential to defense of the study have already been reviewed in this section. Second, such a series of analyses can be meaningfully reviewed only from an operational perspective. That is, before a specific set of findings can be meaningfully evaluated, it is necessary to examine and describe the operations used to measure the dependent variable. Authors such as Campbell and Fiske (1959), Mischel (1968), and Lawson (1965) have clearly shown that the label attached to a set of operations does not mean that it is equivalent to another set of operations with the same label. A detailed operational analysis of the W-K findings is too extensive
to include in this section. Consequently, it seems reasonable to discuss those findings directly relevant to the present study in a later chapter.

In summary, W-K developed a set of operations for measuring creativity which constitute a unitary dimension of cognition development with discriminant validity from general intelligence. The theoretical basis for the W-K operations were largely derived from Mednick's associational theory of creativity. On the basis of the data reviewed, it seems reasonable to conclude that the W-K creativity operations have sound psychometric properties and a promising theoretical base.

The Cronbach Critique of the Wallach-Kogan Study

Cronbach (1968) recently completed a critique of the W-K study as well as a complete reanalysis of their data. Cronbach expressed objection to W-K's presentation of their findings to a general audience in book form when many of their conclusions consisted of speculative hypotheses often supported only by near-significant effects derived from extensive "data splitting." Cronbach did not express opposition to the advancement of speculative hypotheses in exploratory research. In fact, he acknowledged the propriety of such practice. It was, rather, dissemination of such tentative findings to a general audience to which Cronbach objected. Cronbach expressed the opinion that a more conservative standard should be applied to both the data analysis and inter-
pretation if the findings were to be disseminated to a general audience.

Cronbach (pp. 492-93) further criticized W-K's use of the terms "intelligence" and "creativity." He points out that the intelligence index was computed from five STEP subtests which measure achievement, two SCAT subtests which draw heavily on school learning, and three WISC subtests of which two (Block Design and Picture Arrangement) do not correlate highly with total IQ. Thus, Cronbach makes the point that what W-K calls intelligence can probably best be described as school learning. Since the W-K Test lacks validity data, Cronbach stated that the only thing one can conclude about what W-K calls creativity is that it measures number and uniqueness of associations in a game-like setting. Cronbach (p. 493) believes the test can best be described as a measure of responsiveness.

Because of what he believed to be a lack of professional consensus as to the meaning of the two terms, Cronbach stated that neutral names for these two variables should be adopted to avoid interpretations which have not been validated. Consequently, Cronbach adopted the letter A to indicate what W-K call intelligence and the letter F to indicate what they call creativity.

Cronbach (p. 494) also criticized W-K for failing to draw attention to the strength of relationship in the data rather than placing emphasis exclusively on significance
tests. W-K employed an analysis of variance design for the
data in their study. The study was descriptive rather than experimental. In the opinion of Cronbach regression analysis would have been more appropriate. Cronbach employed a step-wise regression procedure in the reanalysis of the W-K data. Since the A (intelligence) variable was the most familiar and best understood, Cronbach began the stepwise procedure by calculating R2 between the A variable and the dependent variables. The F variable and the A x F interaction were then added to the regression equation. The significance of the increment in R2 was tested after the addition of each variable.

Utilization of regression analysis enabled Cronbach to describe significant interactions as a regression function. With a significant interaction, the regression equation assumed the form \( Z = aA + bF + cA \times bF + d \). When the interaction term \((A \times F)\) is not significant, the coefficient for the A x F term is zero and this term drops out of the regression equation.

Cronbach also criticized the failure of W-K to maintain a reasonably high significance level in their hypothesis testing. One of the most serious errors noted was acceptance of mean differences significant at the 0.10 level. Cronbach noted that, "W-K tested around 200 hypotheses regarding effect of A, F, and their interaction; they discuss 42 relations as more or less significant" (p. 494). As Cronbach observed, one would expect approximately half of the 42 tests where significance was obtained to have attained significance by chance
alone. Thus, it is quite probable that a large proportion of the significant effects interpreted by W-K in their discussion were due to chance rather than true differences.

W-K were also criticized by Cronbach (p. 493) for conducting separate data analyses by sex. Cronbach expressed preference for treating the scores of both sexes together except in cases where a significant interaction involving sex is obtained. The rationale for treating the data for both sexes together was justified by Cronbach on the basis of parsimony, i.e., use the simplest correct explanation. In addition, it was noted that within sex analysis reduces the degrees of freedom by half and greatly reduces the power of the significance test. Cronbach (pp. 500-501) discusses applicable regression analysis techniques for determining the presence of a sex interaction.

Finally, Cronbach (p. 493) called attention to the failure of W-K to provide appropriate interpretation for linked scores. For example, these researchers performed separate analyses of variance for number of high, medium, and low quality responses to a given task. Such a procedure frequently provided conflicting results, e.g., when one of the three attained significance and the remainder did not.

In summary, Cronbach's critique indicates that many of the relationships discussed by W-K are relatively weak. The correlates of creativity discussed and interpreted by W-K must be regarded with caution. However, those aspects of the
W-K study criticized by Cronbach are not essential to the present study since neither the dimensionality nor discriminant validity from general intelligence were questioned.

The Wallach-Wing Validation Study

Wallach and Wing (1969) conducted a validation study of the Wallach-Kogan Creativity Test. They employed a sample of 503 incoming freshmen at Duke University on a volunteer basis. This volunteer sample (302 men and 201 women) constituted approximately 40 per cent of the incoming freshman class. Students were informed that their participation in the testing would in no way influence the decision to accept or reject their application for admission to the university. Analysis of mean SAT (verbal and mathematical) scores revealed that the sample was highly similar to the total entering freshman class for that year both within and across sex (p. 30). However, it was found that women scored significantly higher than men in the Verbal subtest of the SAT while men scored significantly higher than women of the Mathematics subtest of the SAT.

The version of the W-K Test employed by Wallach and Wing contained four of the five subtests of the original test. The Instances subtest was not included, perhaps because both scores showed the lowest split-half reliability of the five original subtests. Each of the four subtests employed (Uses, Similarities, Pattern Meanings, and Line Meanings) consisted of three items. The total battery of twelve items was mailed
to the students the summer before their freshman year. A group-test format was employed and the students were informed that participation was voluntary. No time limit was imposed and the participants were urged to take as much time as needed to complete the tasks.

Both number of associations and number of unique associations were scored for each of the four subtests. A unique association was defined as one which occurred only once in the responses of all participants. The four subtest scores for number of associations were each standardized and then summed to provide a total index for number of associations. A similar procedure was followed for number of unique associations. Thus, total scores were obtained for both number and unique associations.

As in the W-K (1965) study, separated analyses were performed by sex. However, analysis for the pooled scores of both sexes were also performed. That is the upper and lower thirds of the sample for each sex were identified on three variables—general intelligence (SAT), total number of associations, and number of unique associations. Thus, scoring high or low on each of the above three measures might be considered the independent variable. High school achievement, freshman grade-point average, and frequency of involvement in various talented accomplishments outside the classroom were cast into the form of dependent variables. The performance of students of each sex scoring in the upper and lower thirds
of the sample on the independent variables was compared on the dependent variables listed above by performance of a t-test for independent means with a directional hypothesis. Mean differences at or beyond the 0.05 level were accepted as significant.

Academic achievement in high school was evaluated by the construction of an index score derived by expressing each student's class-rank in relation to his class size. As one would expect, significant differences were found for men, women, and the pooled scores for both groups. That is, students of both sexes in the upper one-third of their class in intelligence received significantly higher marks in high school than students in the lower third of their class in intelligence.

Students of both sexes scoring in the upper third in number of associations received significantly higher marks in high school than students scoring in the lower third in number of associations. In other words, students who gave the largest number of associations also tended to receive higher high school marks. With regard to number of unique associations, significant mean differences were found for the pooled scores of men and women as well as for the scores of men. However, the mean difference for women alone did not attain significance.

In the case of intelligence, the findings for freshman grade-point average paralleled those for high school
achievement. Again, students of both sexes who scored in the upper third in intelligence also received significantly higher grade-point averages. Significant differences between the high and low association groups were also found on freshman grade-point average for the pooled scores of men and women as well as for men. For women alone, however, the mean differences between the high and low association groups did not attain significance.

With regard to the high and low uniqueness groups, none of the mean differences attained significance. That is, the grade-point averages of men alone, women alone, and both sexes pooled did not differ significantly between the high and low uniqueness groups.

On the basis of the findings reported previously, Wallach and Wing reached several conclusions regarding the relationship between creativity, intelligence and the dependent variables:

Level of academic achievement in college as well as in high school, therefore, provides clear validating evidence for the distinction between high and low intelligence groups in our sample. High intelligence is linked with the achievement of higher grades. Moreover, ideational productivity seems also to make a contribution to grade level in college as it did in high school. This effect furthermore, is not shared with uniqueness of ideas but rather is specific to number of ideas produced. It is as if the student who generates a larger output of ideas on a topic will tend to do so in connection with academic course material as well, with the consequence that he digs somewhat more thoroughly into the content of the course.
But the major empirical message in the present findings is that our hypothesized link between intelligence and average academic accomplishment in college receives very strong confirmation. (p. 63)

Several aspects of talented accomplishment outside the classroom during high school were examined by the authors. A questionnaire was employed which consisted of three to five items covering lack of the following areas: leadership in student organizations, achievement in the visual arts, participation in social service activities, achievement in music and involvement in scientific endeavors. This questionnaire, like the W-K Test, was mailed to incoming freshmen during the summer before they arrived on campus. However, all students receiving the questionnaire had already received notification of their acceptance by the university. Because the questionnaire was mailed at this time and under these conditions, the authors expressed the belief that student responses should provide valid evidence concerning their actual participation in the activities listed above.

In addition, the authors built an argument for the validity of their self-report procedure by an analysis of the percentage of students reporting involvement in each of the activities listed for a given area. Take the area of art as an example. Here 93 per cent of the respondents indicated they had created an art work while only 12 per cent reported having won an award or prize in art competition. Since one would reasonably expect many more students to have created an
art work while only 12 per cent reported having won an award or prize in art competition. Since one would reasonably expect many more students to have created an art product than to have won an art prize or award, the authors reasoned that their procedure was valid. Generally speaking, such analysis of the percentage of students reporting involvement in each of the activities listed for a given area. Take the area of art as an example. Here 93 per cent of the respondents indicated they had created an art work while only 12 per cent reported having won an award or prize in art competition. Since one would reasonably expect many more students to have created an art product than to have won an art prize or award, the authors reasoned that their procedure was valid. Generally speaking, such analysis of responses for each of the seven areas supported the validity of the students' self-reports of extra-school activities during high school.

To determine generality of involvement in all seven areas of talented achievement, Wallach and Wing constructed an index giving equal weight to each of the seven areas. This index was constructed by assigning a weight of zero when no activities were reported, a weight of two for two reported activities, and a weight of three for three or more reported activities. Thus, for each of the activities a student could earn a score of from 0 to 3 making a total score range of from 0 to 21 possible for the seven areas. Separate analyses were performed for students in the high and low thirds on the
following variables: intelligence, number of associations, and number of unique associations.

The results of the analyses noted above largely confirmed the expectations of the authors. No significant mean differences on the breadth of activity index were found for the high and low intelligence groups (men, women, men and women pooled). On the number of associations variable, significant mean differences between the high and low groups were found on the breadth of involvement index for men, women, and the pooled scores for men and women. In the case of the uniqueness variable, significant mean differences between the high and low groups were found for the pooled breadth index scores of men and women as well as for the scores of women alone. The mean differences for men did not attain significance.

These findings led Wallach and Wing to conclude that: "...output of ideas, per se, rather than uniqueness of ideas in particular, is the more potent influence upon the generality of a student's non-academic accomplishment" (p. 78). The authors go on to conclude that an individual's cognitive energy level may well be the critical factor in determining the breadth of his talented accomplishments outside the classroom.

Wallach and Wing also examined the extent of achievement in each of the seven areas of talent previously noted. In each of these areas the dependent variable consisted of the number of items reported by the student. The authors interpreted this figure as an indication of the depth of achievement
for each area.

The first area examined was that of leadership activities reported during high school. Significant mean differences were attained for both men and women singularly and for their pooled score between the high and low association groups. However, none of the mean differences for either uniqueness or intelligence attained significance. These findings indicate that students high in number of associations are involved in more leadership roles in student organizations than students low in number of associations.

In art the second area examined, significant mean differences were found for both number associations and uniqueness of associations. Again, mean differences for the high and low intelligence groups did not attain significance. These findings indicate that students who generate a large number of total associations are the ones who are most involved in art.

In the third area, that of social service, the findings stand in contrast to those previously discussed. No significant mean differences were found on social service involvement for any of the three independent variables. That is, students high in total associations and in uniqueness of associations did not show greater social service involvement than students low in total associations and in uniqueness of associations. Again, the high and low intelligence groups did not differ with
regard to social service involvement.

In writing, the fourth area, the achievements of the high and low intelligence groups of both sexes failed to show significant mean differences. In other words, students groups achieved at about the same levels in extraclass writing activities. Significant mean differences in writing achievement were found for number of associations in the case of women and for the pooled scores of men and women. The mean difference for the high and low association groups of men did not attain significance. However, significant mean differences were found for the high and low uniqueness in the case of both men and women as well as for their pooled scores. Wallach and Wing (p. 91) summarize their findings as follows:

The contrast that we have found in the writing domain between the strong positive linkages with ideational output and uniqueness, on the one hand, and the lack of any connection with intelligence level, or the other, seems very impressive indeed.

In the fifth and sixth areas, dramatic art and music, none of the high-low group contrast attained significance. The findings for intelligence confirmed the expectations of the authors. The authors advanced the hypothesis that both dramatic and music performance involve skill at the reproduction of material already written.

Since both the music and dramatics skills topped by the questionnaire involved performance rather than production, according to Wallach and Wing's interpretation, one would not expect either associative production or uniqueness to exercise
a significant effect. If either composition or play writing were involved one would have expected different results.

With regard to extra-school science attainment, the seventh area, mean differences between high and low intelligence groups did not attain significance for men, women or the pooled scores of men and women. For both men and women and their pooled scores, significant high-low group differences in science attainment were obtained on the number of associations variable. That is, students who gave the largest number of associations also showed the greatest depth of involvement in extra-school science activities. For women as well as for the pooled of men and women, significant mean differences were found between the high-low contrast groups on the uniqueness dimension. Mean differences for men did not attain significance on this variable.

The results reported by Wallach and Wing for the seven areas appear to present a clear pattern of findings for both intelligence and ideational productivity (number of associations). No significant mean differences between the high and low intelligence groups were obtained in any of the seven areas. Significant mean differences for the high and low ideational productivity groups (sexes pooled) were obtained in four areas: leadership, art, writing and science. In the case of ideational uniqueness (number of unique associations) the results for the two sexes pooled paralleled those for ideational productivity with the exception of the leadership area where the mean differences
did not attain significance.

In addition to the analyses for the seven areas discussed above, Wallach and Wing performed a chi-square test for the frequency of response to all of the items within the seven areas. That is, a Chi-square test was performed to determine whether the students in the high and low intelligence groups responded differently. Parallel analyses were performed for students in the high and low groups on ideational productivity (number of associations) and ideational uniqueness (number of unique associations).

The four areas in which overall significant differences were obtained (leadership, art, writing, and science) for ideational productivity contained twenty items. Fifteen of the 20 ideational productivity (high-low) comparisons yielded significant results while only 5 of the 20 ideational uniqueness (high-low) yielded significant results. In contrast, only one of the 20 intelligence comparisons yielded significant results. Wallach-Wing interpreted these findings to support their position that ideational production represents a kind of cognitive energy which is generalized across a wide spectrum of activities. The following quotation appears to reflect the present position of Wallach and his associates:

These findings clearly point to the conclusion that level of ideational output in general, rather than producing unique ideas in particular, is what matters most for the non-academic attainments under investigation. (p.110)
Further Research Relevant to the Wallach-Kogan Study

A recent study by Cropley (1968) also provides information concerning the validity of some of the constructs advanced by Wallach and his associates. The complete creativity battery was administered to a volunteer sample of 124 freshmen at an Australian university. A group format was employed in which the student wrote his responses to the test on a record form. An attempt was made to maintain freedom from time pressure and an informal atmosphere as recommended by W-K (1965). In addition to the creativity battery, five conventional measures of intellectual ability and achievement were included. The creativity battery was scored for both number and uniqueness of associations. For purposes of data analysis, raw scores from the tests of ability and intelligence were employed.

The data analysis revealed findings which largely support the formulations of W-K, intercorrelations among the ten creativity subtests were generally substantial. At the same time, scores on the creativity subtests proved to be independent of general intelligence. Yet, the creativity subtests did not intercorrelate so highly as the various scores obtained from the intelligence and achievement tests. This was especially true for the uniqueness scores.

A principle axis factor analysis (with unities in the diagonals) was performed on the 15 x 15 intercorrelation matrix. This analysis revealed the existence of a very strong
general factor which accounted for about 41 per cent of the common variance among the 15 variables along with a second strong factor which accounted for approximately 30 per cent of the variance. The second factor was interpreted as a "...substantial bipolar factor of creativity versus intelligence" (p. 200). The intelligence and achievement variables loaded very heavily on this second factor while all the creativity subtests loaded negatively but to a lesser degree. All variables showed substantial to very strong positive loadings on the general factor. These results led Cropley (pp. 200-201) to draw the following conclusion:

Thus, provided that the presence of a considerable general factor is kept in mind, the Wallach-Kogan tests can usefully be used in the sampling of creativity, even with age levels other than that of grade five children, and in a group rather than an individual form.

Ward (1967) reported a factor analysis of the intercorrelations from the original W-K (1965) data. Ward performed a principle components analysis in which four significant factors were obtained. These factors were rotated to Varimax criterion. The Promax method was employed to obtain an oblique solution for the Varimax factors. This procedure provides a factor pattern, a factor structure and the correlation between the oblique factors. A second-order oblique analysis of the primary factors was also performed. However, the results of this analysis were not reported because of a lack of space devoted to the article.

Ward's interpretations of the four primary factors
is most interesting. Factor one was interpreted as a measure of school achievement. Factor two was a creativity factor on which all ten of the subtest scores loaded positively. Factor three was so poorly defined that interpretation is almost impossible although Ward felt that it represented scoring for number of responses as opposed to uniqueness and accuracy of response. Finally, factor four was interpreted as either a weak general factor or a measure of social atmosphere. Ward (p. 382) summarized the implications of his findings as follows:

It was not originally intended to prove or disprove the inferences drawn by Wallach and Kogan, although the results tend to support their choice of procedure. However, the multifactorial nature of "creativity" data is once more demonstrated despite the presence of two apparently near orthogonal and easily identifiable sets of measures.

Fee (1968) also conducted an independent factor analysis of the W-K (1965) data. The centroid method of factor extraction (with unities in the diagonals) was used to extract first-order factors. Horst's multiple group factor method was employed for deriving second-order factors and transferring them back to the first order realm through rescaling. Four first-order primary factors and two rescaled first-order factors corresponded to Ward's bipolar creativity versus intelligence factor. Fee's analysis revealed two primary creativity factors, one visual and one verbal. Another primary factor clearly associated with school achievement was
obtained. A fourth primary factor of non-verbal intelligence was identified by its loadings on Picture Arrangement and Block Design subtests from the WISC.

Despite methodological disagreements, the Ward and Fee studies agreed in showing that the W-K Test measures a dimension relatively distinct from conventional intelligence tests. At the same time, both authors also agreed that creativity is probably not a unidimensional phenomenon. This would mean that interrelationships with variables not treated by W-K will eventually have to be considered.

Elkin, Deblinger, and Adler (1970) reported a study investigating the sensitivity of the W-K test to motivational influences. Items from the Instances, Similarities, and Alternative Uses were subdivided on an odd-even basis to provide equivalent forms of the test. The main independent variable consisted of a self-selected activity on the part of the child such as special interest work in science, reading, music, shop, or drama. The uninteresting activity consisted of a boring, routine task such as circling all n's and 6's on a page. In both cases the children knew they would be returning to the interrupted activity after finishing the W-K test which was presented as a game.

A repeated measure designed with counterbalancing of treatment (type of activity interrupted) was employed. This design enabled the investigators to determine the effects of order-of-motivating condition and groups-under-order-of-motivating-condition as well as motivating condition which was the indepen-
dent variable of major concern. Data analysis revealed a strong effect (p .01) for motivating condition. That is, pupils removed from an uninteresting condition to which they had to return scored much higher on the W-K test than pupils removed from an interesting activity to which they had to return.

Research and Theory Concerning Criterion Development

In their discussion of creativity criteria, Cattell and Butcher (1968) made the statement, "Seldom has psychology been asked to undertake so ambitious a task as that of defining the creativity criterion" (p. 285). These authors pointed out, rather succinctly, that little can be accomplished in creativity research until psychologists are willing to expend the effort necessary for the development of suitable creativity criteria. After all, these researchers note, if psychologists encounter difficulty in developing a reliable criterion for successful bus driving, how much more formidable is the challenge faced in developing effective creativity criteria?

At a later point in their discussion, Cattell and Butcher (p. 301) expressed a preference for concentrating on criterion development at the adult level, at least for the present time. Substantial concerns prompted this position. First, these authors felt that creativity criteria are extremely difficult to develop and evaluate for children's products. Second, they noted the lack of certainty as to whether creative child-
ren grow up to be the most creative adults since the restrictions and demands of adult life are different from those of childhood and adolescence. Finally, Cattell and Butcher maintained that since creative performance in adult life is the ultimate concern of education, primary emphasis should be placed at this level.

Thinking specifically in terms of scientific creativity, Cattell and Butcher advanced several other fruitful ideas concerning criterion development at the adult level. Two criteria were proposed. The first of these was a quality rating of originality and significance of the scientist's work by eminent colleagues. When applicable, number of patents would also be included as part of this quality criterion. The second proposed criterion consisted of quantity of research articles accepted by reputable journals. The product of these quality and quantity criteria, Cattell and Butcher believe, should constitute an appropriate creativity index for use by researchers desiring a single creativity criterion.

After outlining what, in their opinion, constitutes the most desirable approach to criterion development, Cattell and Butcher went on to discuss the most serious sources of error in the implementation of their proposed procedure along with possible procedures to ameliorate the situation. The first potential source of systematic error in implementing the criterion development procedures outlined above stems from attempting to evaluate "...what belongs to the future by concepts that be-
long to the present and the past ... (p. 288). A second potential source of systematic error stems from "... the over-evaluation of what can be understood by the many at the expense of what can only be understood by the few" (ibid.). To overcome these sources of error, Cattell and Butcher suggested periodic evaluation, at ten year intervals, of scientific contributions in various fields of endeavor in order to determine which elements agree best with future evaluations. Secondly, these authors suggested that the popularity effect be controlled through determining what discoveries decline in importance as one proceeds from the ratings for average members of a profession through the most eminent ones.

Cattell and Butcher discussed two methodological approaches to criterion development for the field of scientific creativity. In the first instance, an adjustment criterion is employed; an effectiveness criterion is used in the second case. Both of these criteria have been frequently employed in criterion based studies. With adjustment criteria, the emphasis is upon determining similarities and differences between groups of individuals in different professions or of different levels of creativity. An example of a study employing an adjustment criterion was a comparison of eminent physicists, psychologists, and biologists by Cattell and Drevdahl (1955). Essentially, studies in which adjustment criteria are employed call for between-group comparisons. When an effectiveness criterion is employed, one measures level of performance in some field and then determines
regression coefficients for personality, ability, and motivational factors upon this effectiveness criterion.

Brogden and Sprecher (1964) also presented an illuminating discussion of criterion development which treats aspects of the problem not previously mentioned in this review. The focus of the article by Brogden and Sprecher, like that of Cattell and Butcher (1968), tends toward scientific creativity at the adult level. In addition, this discussion tends toward methodological considerations, as opposed to theoretical considerations, in criterion development. Throughout their discussion, however, these authors made very clear their conviction that progress cannot occur in creativity research unless more emphasis is given to criterion development.

Among the most fruitful aspects of this discussion was the distinction between process and product criteria. Process criteria were defined as behaviors necessary for the achievement of creative products. An example of a process criterion might be grace, poise, or coordination displayed in the completion of a motor task. Product criteria, in contrast to process, have existence apart from the person involved in the production. Examples of product criteria might be a theoretical system, an equation, a painting, or a musical composition.

Both authors agreed that product criteria are the most vital for creativity research noting that "...there is little question that the approach through the product will most closely approximate the ultimate criterion" (p. 158). Yet, there was
some disagreement between the authors as to the value of process criteria with one feeling process criteria are not of value while the other felt that process factors may be involved in ultimate creativity criteria.

Both Brogden and Sprecher agreed that, "Despite the fact that products lie at the heart of criterion problems, little work has been done with them" (p. 158). Several relevant suggestions were advanced for the development of product criteria. First, it was suggested that a distinction be made between the absolute and relative value of a product. In the case of the latter, the value of a product is evaluated in reference to input costs as well as to judged merit. Second, it was suggested that a distinction be made between products that occur in the normal course of events and those that are produced on request for creativity research, i.e., work-sample measures. Third, it was suggested that the relationship between amount of creative productivity and level of creativity attained by individuals be explored. The related question of the relationship between level of creativity and diversity of creative products was also suggested for research. Fourth, it was suggested that the factors which are responsible for a given judgement should be studied. Alternatively stated, what is the source of the value judgement made by the rater about the product? Fifth, research concerning the influence of rater qualifications was recommended. For example, do professional artists rate children's art products differently than elementary art teachers?
In addition, Brogden and Sprecher discussed possible methodological approaches to the problem of overlap among creativity criteria. Two approaches were suggested for simplifying this problem. The authors suggested that factor analysis, one of the recommended approaches, can be employed in two ways. First, factor analysis can be used in the conventional manner to simplify the dimensions of either a predictor or criterion matrix. Second, a factor analysis of the predictor variables can be performed with criterion variables added later to the results to check on their equivalence.

The alternative to the factor analytic approach is the criterion-equivalence approach. With this approach it is necessary that the same predictor variables be available for both criteria. A multiple-regression equation is developed for the first criterion. The same procedure is followed for the second criterion using a predictor matrix containing the same variables. To the extent that the multiple-regression equations for the two criteria are similar, the content of the two criteria are also similar. Comparison of the respective beta-weights for the two criteria often provides insight as to specific similarities and differences in the content of these criteria.

Furthermore, the discussion by Brogden and Sprecher provided a synopsis of the major suggestions for criterion development which evolved in the course of the Utah creativity conferences. Since the majority of the original Utah conference reports are out of print and unavailable, it was often necessary
to rely on this secondary source for information. Among the suggestions from the earlier conferences was that the heaviest weight be placed on an individual's best contribution in criterion development. The rationale for this decision was that a person's highest achievement is the best indicator of what he is capable of doing.

Another suggestion was that professionals of proven ability and reputation should be employed as judges of creative products in their areas of competence. Also, it was suggested that examples of products at different levels of creativity be collected. For example, one might collect examples of children's crayon drawings in each of nine categories ranging from superior through average to very poor. These products could then be used as standards or reference points in future creativity research for judging the drawings of children from the same age group. The criterion considerations noted above are far from exhaustive in terms of the total list enumerated by Brogden and Sprecher from the Utah conferences. However, the concerns discussed above were the ones of direct relevance to the criterion problems of concern in the present study.

Finally, Brogden and Sprecher discussed variables which investigators should attempt to control in their creativity research. Three categories of control variables were discussed. These were biographical, time, and opportunity variables. Biographical variables refer to background factors such as early home life, educational history, and employment history.
Time variables refer to the influence of different ages at which an individual is studied. Opportunity variables refer to such factors as material resources, the psychological and social environment, and the difficulty of the problem area worked on.

The work of Calvin Taylor and his associate, e.g., Taylor, Smith and Ghiselin (1963) and Taylor and Ellison (1964) probably represents the most thorough and comprehensive criterion research yet conducted in the area of creativity. Their samples were, however, limited to the physical scientists and engineers working at two large government research installations. In terms of methodology, Taylor employed both factor analysis and the criterion equivalence approach. That is, factor analysis was performed to determine the factor structure of the criteria. Multiple regression equations were then determined for each criterion factor, and then cross-validation was conducted with another sample. It was found that selected biographical variables held up very well in cross-validation.

Stated briefly, the most important single finding of Taylor's research in criterion development is that creativity criteria are extremely complex and multi-dimensional. As a consequence, Taylor and his associates have frequently called attention to the need for multiple criteria in crea-
tivity research. Secondly, biographical variables proved to be the best type of predictors for the creativity criteria. The biographical data were collected by use of an inventory completed by the scientists in the sample. A modification of this inventory was successfully employed in predicting scientific creativity criteria for high school students attending a summer workshop.

The research reported by Beittel (1964) provides considerable encouragement because of the success encountered in developing creativity criteria. Beittel and his associates have been intensely concerned with developing creativity criteria in the visual arts. A process criterion, deliberateness versus spontaneity, was identified and carefully studied by this group. This process variable is rated from the subject's finished product. That is, the process is inferred from the product.

The spontaneity end of this dimension refers to the extent to which an individual works with spontaneity, feeling, and relaxation as opposed to determination, deliberateness, rigidity, and tightness. High interjudge agreement has been consistently obtained for products rated on this dimension. The spontaneity-deliberateness dimension as well as several other criteria have been studied by Beittel and his students. One of the most outstanding characteristics of the work of this group has been the effective combination of predictive and experimental research.
CHAPTER III

RATIONALE FOR THE PRESENT STUDY

There is general consensus that the creativity researcher faces formidable obstacles in the design and implementation of his research. When children are employed as subjects the obstacles are even more forbidding. There are two major reasons for the lack of knowledge. First, relatively few researchers have focused their attention on creativity research with children. A related problem also stems from the fact that, with a few notable exceptions, creativity research with children has not maintained the interest of the majority of the individuals who have published in the area at one time or another. In other words, creativity research with children might be characterized as an area which is casually approached and casually discarded as an area of major research interest. Such lack of sustained interest causes a discontinuity in the flow of research which is not conducive to the advancement of knowledge.

Second, the majority of the existent creativity research with children has been superficial in nature and lacking in methodological sophistication. A good example of poor planning and inappropriate methodology is to be found in the time and effort wasted in attempting to determine the relation-
ship between intelligence and creativity as defined by the Guilford-type divergent production test. This question could have been settled in the course of test construction as Flescher (1963) did or at most one or two well designed studies could have resolved the issue.

One of the most important aspects of the lack of methodological sophistication has been the neglect of the criterion problem. With the exception of the Martinson and Seagoe (1967) study one is at a loss to locate a study where creativity operations have been used to predict the quality of children's creativity products. This neglect is especially disconcerting when the need for criterion development was pointed out during the early stages of creativity research by individuals such as Brogden and Sprecher (1964), Taylor (1964), Astin (1964), Beittel (1964), Taylor and Holland (1964), Ghiselin (1963), and Harmon (1963). Furthermore, several publications noted above are comprehensive enough to provide the creativity researcher with a great deal of assistance in criterion development.

The lack of criterion validity for the most widely used creativity tests for children constitutes a serious constraint to advances in creativity research. Underwood (1957) described the relationship between the process of inquiry and the advance of knowledge. Knowledge advances through the research process. Ideally, the advance of knowledge leads to an increased congruence between theory and the operations derived from the theory under consideration. Concomitantly, data
obtained through application of relevant operations result in modification of the theory from which the operations were originally derived.

Unfortunately, this process has not developed where creativity research with children is involved. The lack of criterion data is the main constraint to the application of the above process to creativity research. In short, there is a dearth of data concerning the relationship between creativity tests and creativity as evidenced in products resulting from relevant activities in the real world.

The remainder of this chapter is devoted to a consideration of the problems discussed above. First, attention is devoted to the problem of selecting the most promising set of creativity operations for use in the study. Second, a rationale is developed for the criteria employed in the study. Third, a brief treatment of design is undertaken to justify certain specific procedures undertaken in the study. In specific, the use of multiple regression analysis and factor analysis are discussed and a rationale for the sampling procedure is developed.

Selection of a Creativity Instrument

Three major criteria were established for the selection of a set of creativity operations. First, this creativity operation under consideration must measure a unitary dimension of cognitive behavior with discriminant validity
from general intelligence. Operationally speaking, this criterion implies that the subtests of the creativity battery must correlate significantly—and substantially—among themselves while, at the same time, showing negligible correlations with conventional measures of general intelligence. Second, the creativity operations under consideration must have demonstrated a satisfactory level of reliability. Third, the creativity operations under consideration must be based upon a substantial body of theory which provides construct validity for the operations under consideration.

The first criterion stated above is of critical importance for creativity research. With regard to this issue, the salient point for the present discussion is that unless it can be shown that a set of creativity operations defines a unitary from general intelligence, one is not justified in discussing creativity in the same way one discusses general intelligence, i.e. as a unitary, pervasive psychological dimension. Numerous eminent creativity researchers and methodologist, e.g., Burt (1962), Thorndike (1963), Marsh (1964), Vernon (1964), and W-K (1965) have pointed out that it is imprudent to sum the scores from creativity subtests to obtain a creativity index unless this first criterion is satisfied.

A review of the literature in search of creativity operations suitable for use with children revealed two alternatives. These alternatives are the creativity battery devised by W-K (1965) or the divergent production tests of
Guiford (e.g. 1967) or some adaption thereof such as found in the work of Getzels and Jackson (1962), Flescher (1963), or Torrance (1966). Therefore, a comparison of these two approaches to creativity instrumentation is undertaken for each of the three criteria outlined above.

A comparison of these two approaches for the first criterion revealed substantial differences. The creativity operations of W-K (1965) were designed to measure a unitary dimension of cognitive behavior divergent from general intelligence. All research to date has shown that the operations developed by W-K are in fact independent of general intelligence and also define a unitary cognitive dimension with substantial common variance in all subtests. In contrast, the divergent production tests of Guiford (1967) Torrance (1966), Flescher (1963), Getzels and Jackson (1962), and others have failed to define a cognitive dimension with sufficient common variance to warrant summing subtest scores to obtain a meaningful creativity index.

Methodologists such as Burt (1962), Thorndike (1963), Marsh (1964), and Vernon (1964) have shown that the variance common to the subtests of the above divergent production tests have consistently failed to show divergent validity from general intelligence as described by Campbell and Fiske (1959). In brief, it can be readily seen that the creativity operations developed by W-K (1965) satisfied the requirements of the first criterion while the divergent production tests considered failed
to satisfy this criterion.

In passing, it should be noted that Mischel (1968), pp. 89-90), one of the most notable critics of trait theory, cites the good judgement of W-K is establishing divergent validity from general intelligence for their creativity tests. At the same time, this author criticized the failure of researchers using the divergent production tests for failing to attain either divergent validity or undimensionality for their instruments.

With regard to the second criterion, adequate reliability, both the divergent production tests based on Guilford's work and the operations developed by Wallach and his associates appear to be satisfactory. Since both sets of creativity operations attained acceptable levels of reliability, it is impossible to discriminate between them on the basis of this criterion.

The third criterion established for the selection of a creativity instrument was that of construct validity. That is, they follow from a theoretical statement as to the nature of creativity. While this criterion was given consideration, it was secondary to the three major criteria previously discussed because of the operational nature of the study. That is to say, the criterion of construct validity was considered only after the first three criteria were satisfied.

In any case, there is a substantial body of theoretical support for the W-K test. The theoretical basis for this test was originally based on Mednick's (1962) associational theory of creativity. In this regard, it should be mentioned
that the creativity battery developed by W-K follows more
directly from Mednick's creativity theory than his own
Remote Associate Test. This is true since the basic units
of Mednick's response gradient was number and uniqueness
of associations. W-K made direct use of these units in the
development of their battery while Mednick required that
several associations be elicited conjointly by a set of stim-
uli to produce a single response specified as correct by the
author. Wallach and his colleagues have elaborated the theo-
retical basis of their creativity test in the course of their
research activities (W-K, 1965), Wallach, (1967), and (Wallach
and Wing, 1969).

Thus, it can be seen that the W-K Test has a sub-
stantial theoretical base which provides a strong basis for
construct validity. However, the divergent production tests
of Guilford (1967) also have a strong theoretical base since
they constituted an integral part of the structure-of-intellect
model. By way of contrast it can be seen that Guilford's di-
vergent production tests are a subtest of a grand theory of in-
telligence while the associational test of Wallach and his col-
leagues is based upon a theory more limited in scope, i.e. which
is limited to creativity. Thus, both the associative battery
of W-K and the divergent production tests of Guilford are de-
rived from promising theoretical formulations which will proba-
bly be extended and polished by their respective authors.
The same case, however, cannot be built for the test developed by Flescher (1963), Getzels and Jackson (1962), and Torrance (1966) which are based on Guilford's concept of divergent production. These tests do not necessarily represent accurate extrapolations of the structure-of-intellect model. As such, these instruments measure what the respective authors believe constitutes the essence of creativity. With the possible exception of Torrance (1966), these authors have not taken the trouble to present a thorough statement of their theoretical views concerning creativity.

In summary, the W-K Test was selected because it was the only creativity instrument available which defined a unitary dimension of cognitive behavior independent of general intelligence. Furthermore, this battery has demonstrated a satisfactory level of internal consistency reliability. Finally, the work of Mednick (1962), W-K (1965), Wallach (1967), Wallach and Wing (1969), and Wallach (1970) provides a promising theoretical base which provides construct validity for this creativity instrument.

Selection of the Creativity Criteria

The literature review presented in the previous chapter revealed only one precedent for developing creativity criteria for children's creative products. Martinson and Seagoe (1967) reported a study in which eight separated creativity criteria were employed for each child. These eight criteria were: (1) recorded answers to science questions, (2) creative
writing of prose, (3) creative writing of poetry, (4-5) written solutions to two social studies problems, (6) clay products, (7) calcimide drawings, and (8) films of interpretive rhythms.

Each of these criteria was rated separately for originality and effectiveness of expression by three independent judges who employed a nine point scale to record their evaluations. That is, for each criterion three subject matter specials in that respective area made independent evaluations of the children's products. Each judge's rating of originality and effectiveness of expression was averaged to obtain a composite score for each pupil in a given criterion area. Then these mean composite ratings from each of the three judges were averaged for obtain a final composite score for each pupil in each respective criterion area.

Average correlations among judges' ratings varied considerably for the eight creativity criteria. The average product-moment correlation coefficient for the four areas of highest interjudge agreement were as follows: clay products, 0.95; interpretation of rhythms, 0.83; science problems, 0.52; and calcimine paintings, 0.45. In contrast, the interjudge reliabilities for the four areas of least agreement were as follows: Utopia, 0.43; Desert Island, 0.38; creative writing—prose, 0.23; and creative writing—poem, 0.04.

The Martinson and Seago (1967) exercised a substantial influence on the present study by demonstrating that inter-
judge agreement can be secured on creativity ratings of children's products in selected areas. Furthermore, this study also provided a set of promising rating scales for the evaluation of originality and effectiveness of expression of children's products. While this study can be criticized on several methodological issues, it cannot be denied that the study represents a fruitful beginning in an extremely difficult and complex area where, as Cattell and Butcher (1968) note, it is easier to criticize than to suggest constructive alternatives.

The research reviewed in the previous chapter suggests several minor modifications in the approach of Martinson and Seagoe. The methodological approach employed by these authors was, roughly speaking, similar to the adjustment criteria discussed by Cattell and Butcher (1963) in that two groups differing on some characteristic were compared. In the opinion of the writer, what Cattell and Butcher (p. 295) describe as regression on a criterion of effectiveness within a group is more appropriate than an adjustment criterion since the former permits the use of powerful regression analysis techniques which provide a measure of strength of association.

Attention was also given to what Brogden and Sprecher (1964) refer to as control variables, i.e., biographical, time, and opportunity variables. The opportunity variable was controlled by eliminating all pupils from the sample who had received special instruction in visual art. Furthermore, a large majority of the
pupils, approximately 4/5, in the sample had attended the same school and consequently received similar art instruction. These precautionary measures should, to a large degree, provide sufficient control for opportunity variables. With regard to time variables, chronological age was made a control variable so that its influence could be evaluated. That is, chronological age was made one of the predictor variables.

Special care was necessary in the case of biographical variables since the research reported by Taylor and Ellison (1964) and Taylor, Smith, and Ghiselin (1963) indicated that some combination of these variables may well represent the most effective set of predictors for creativity criteria. The findings of these studies noted above were derived only from samples of adults and older adolescents. However, enough variables from early home and family background contributed sufficient variance to the final $R^2$ to justify careful control. The following biographical variables were included in the design of the present study for purposes of control; intactness of home, education of parents, religion of parents, race, sex of pupil, and birth order.

Martinson and Seagoe (1967) did not always specify whether they collected their products in the course of normal instructional activities or on a work sample basis. However, it would appear that in at least some cases special conditions not typically found in a regular classroom setting were present. If this inference is justified, then at least
some element of a work sample situation was introduced. Collection of work sample data often produces a test-like atmosphere which may differ substantially from that found in a regular art session conducted in a self-contained classroom. To ameliorate this situation the art products were collected by the regular classroom teacher during periods ordinarily scheduled for art instruction. In this way standardized data collection procedures were introduced into regular instructional activities without the pupils being aware of any deviation from regular instruction. This approach is in accordance with the procedure recommended by Brogden and Sprecher (1964, p. 163).

The two areas selected for criterion development were of no special significance since the constructs basic to the W-K test describe a broad dimension of cognitive behavior commensurate with general intelligence. In other words, this battery should effectively predict creative performance in any area. Clay products were selected because the Martinson and Seagoe (1967) study indicated that excellent interjudge agreement could be secured for product ratings in this area. Crayon drawings were chosen because a second sample of work in the visual arts was desired. In addition, the pupils had a great deal of experience and familiarity with the media selected.

Considerations in the Design of the Study

An extensive review of the literature indicated that
multiple regression analysis represents the most effective methodological approach for the present study. In a recent exposition Cronbach (1969) illustrated the advantages of employing regression analysis for the study of descriptive data. With regression analysis both a measure of relationship and a test of significance are obtained. In contrast, application of complex analysis of variance provides only a test of significance. With analysis of variance, a significant difference indicates only that it is highly improbable that the obtained results are due to chance. Thus, the researcher knows that a given variable exercises some influence on the dependent variable but he is without guidance in determining whether this influence is large enough to be of interest. For example, if Martinson and Seagoe had employed regression analysis it would have been possible for them to determine the proportion of criterion variance accounted for by each of the independent variables.

Factor analysis is employed to determine the factorial composition of the predictor and criterion variables. That is, a factor analysis was performed to provide additional information for understanding the multiple regression equations obtained for the creativity criteria. A knowledge of the hierarchical factor structure of the predictor and criterion variables permits the researcher to examine the structural relationships among the basic dimensions involved in the intercorrelation matrix.
Three major considerations guided the selection of the sample. First, it was decided that the sample should differ from the one selected by W-K in that pupils with a different socio-economic, ethnic, and racial background should be included. Second, it was necessary to locate a school which had adequate facilities for three psychologists to conduct evaluations at the same time during a specified block of time. Third, it was necessary to secure the cooperation of school officials for conduct of the study. Given these considerations, convenience played an important role in the selection of the sample.

However, the fact that the sample is not necessarily representative of any particular subset of "so-called" disadvantaged pupils does not constitute a serious constraint because the research previously discussed shows that creativity, as defined in this study, constitutes a unitary cognitive dimension independent of general intelligence which is present to some degree in all children. The concern of the study is with creativity as a monotheic cognitive dimension. Consequently, the W-K Test should predict creative performance on the criterion variables across any reasonably broad subset of the creativity continuum.

Selecting a sample which is relatively homogeneous with regard to biographical and situational variables has the effect of increasing the probability of a type II error, that is, of failing to obtain correlations between predictors and
criteria as large as one would find in a more heterogeneous sample. From this standpoint the present study represents a conservative approach to creativity research.

In any case, the present study also represents a very practical approach to the validation of the W-K Test because educational and psychological tests are most frequently used with class-size groups. Consequently, the failure of this battery to maintain its validity for any class of pupils in the average range of intelligence could result in serious consequences.

Since convenience played a substantial role in the selection of the sample, it was decided that the most desirable approach would be to describe the characteristics of the sample in detail. This procedure allows the reader to examine the status of the sample on any variables of interest to him and draw his own conclusions regarding the generalizability of the findings. The use of indices was avoided in the description of the sample. This procedure is in accordance with the advice of Coleman (1964, pp. 75-84), an eminent mathematical sociologist.
CHAPTER IV

METHOD AND PROCEDURE

An important consideration in the preparation of the design of the present study was the assumption that the careful collection of a large amount of data concerning each child through individual assessment procedures is preferable to the collection of less comprehensive group data for a larger sample. The procedures employed in this study were chosen to investigate the relationship between children's creative products in selected areas of art, the criteria, and variables hypothesized to correlate with creative performance in art. The specific variables of concern were: creativity test scores, individual intelligence test scores, and judges' ratings of originality and effectiveness of expression for children's clay and crayon products. In addition, nine control variables were incorporated in the manner recommended by Brogden and Sprecher (1964). These variables were incorporated into a design which should provide information concerning the validity of the W-K Creativity Test for the prediction of concurrent creative performance in two areas of visual art.
General Hypothesis

The general hypothesis of the study was that the operations developed by W-K (1965) for measuring creativity will prove to be effective predictors of performance on the creativity criteria. This hypothesis was based on an examination of the conceptual basis of W-K's creativity operations as well as a review of relevant research. The W-K operations have not been validated against criteria in the visual arts. However, these authors have developed a generalized theory of creativity which is supported by a substantial body of experimental, descriptive, and factor analysis research. However, the overall trend identified by Martinson and Seagoe (1967) suggested that general intelligence may also be an effective predictor of creative performance across a wide variety of tasks. Furthermore, the research of Calvin Taylor and his associates (e.g. Taylor and Ellison, 1964) strongly suggests that biographical variables are the most effective predictors of creativity criteria—at least at the adult and adolescent levels.

Thus, it can be seen that three classes of predictor variables were strongly suggested in the literature. With these findings as a precedent, it was decided that variables from all three of these classes should be included in an attempt to determine the most effective combination of variables for predicting the creativity criteria. Such an approach is congruent with Underwood's (1957) analytic approach
as well as the recommendations of Brogden and Sprecher (1964) and the strategy employed by Taylor and Ellison (1964). That is, the formulation of specific null hypotheses were avoided because commitment to such a priori hypotheses often commit one to a series of meaningless exercises when he faces an area in which there is a dearth of knowledge. This occurs because one is not in a position to delineate the range of potential outcomes. In accordance with this position the following questions were formulated:

1. To what extent is the discriminant validity from general intelligence and high subtest intercorrelations reported by W-K and others generalized to pupils in the present sample with a different social, ethical, and economic background?

2. Will the four judges achieve a significant degree of concordance in their ratings of the artistic products submitted to them for judgement?

3. What is the factorial composition of the predictor and criterion variables employed in the study?

4. Among the variables examined; five creativity subtests, eleven WISC subtests, and nine control variables; what is the most effective combination for predicting concurrent creative performance in the areas of visual art selected for study?
Sample

The sample employed in this study consisted of 73 third and fourth grade pupils in an inner-city parochial elementary school of Columbus, Ohio. With the exception of two pupils dropped from the sample because of special training in art, the sample included all pupils enrolled in the third and fourth grades of this school. The grade-by-sex breakdown showed 45 girls, 22 third graders and 23 fourth graders, and 28 boys, 11 third graders and 17 fourth graders.

The rationale for the sampling procedure was developed in the previous chapter. Consequently, the remainder of this section is devoted to a description of pupil background characteristics of possible relevance to creativity. Consideration of several variables seemed to be in order.

First, consideration of the general intellectual level of pupils in the sample seemed to be in order. Analyses of this variable revealed an average WISC (full scale) IQ of 100.47 with a standard deviation of 11.34. These data indicate that while pupils comprising the sample are very typical of the general population in average ability, they are somewhat more homogeneous than the general population of pupils their own age. Appendix A contains a frequency distribution of the intelligence quotients for pupils in the sample.

Second, the educational level of the parents was considered. Analysis of this variable revealed that the years of formal education completed by the fathers were as follows:
college graduates, 5; some education beyond high school, 3; high school graduates, 38; some high school, 15; eighth grade, 7; less than eighth grade, 12; and no information available, 2. Thus, it can be seen that the majority (38/73) of the fathers were high school graduates, a few (8/73) were college graduates or had some college, and a substantial number (29/73) did not complete high school.

The years of formal education for mothers was as follows: college graduates, 2; some education beyond high school, 6; high school graduates, 41; some high school, 15; eighth grade, 7; and less than eighth grade, 2. Thus, the majority, 41/73, of the mothers were high school graduates; very few, 2/73, were college graduates; a few, 6/73, had some training beyond high school; several, 15/73, had some high school; and some, 9/73, had an eighth grade education or less.

Third, the occupational status of fathers of pupils in the sample was considered. The classification system of the Dictionary of Occupational Titles, Vol. II. was employed to describe the occupational status of the fathers (U.S. Dept. of Labor, 1965). According to this system the occupational status of the fathers was as follows: professional, technical, and managerial, 10.9 per cent (8/73); clerical and sales, 13.7 per cent (10/73); service occupations, 21.9 per cent (16/73); skilled laborers, 13.7 per cent (10/73); semi-skilled laborers, 6.8 per cent (5/73); unskilled laborers, 6.8 per cent (5/73); unskilled laborers, 20.6 per cent (17/73); and unemployed and/or
occupational status unknown 12.3 per cent (9/73).

Fourth, the birth order of pupils in the sample was considered. The pupils in the sample range from the first through eighth in birth order. Examination of the distribution of birth order readily revealed that the pupils in the sample were part of much larger than average families. This statement can be readily verified by examination of Appendix B which contains a frequency distribution for the birth order variable.

Fifth, analysis of the chronological age variable was also in order since the sample was drawn from two different grades, thus resulting in a rather wide age range. The chronological age of pupils in the sample varied from 102 through 143 months. The mean age for the sample was 116.73 months with a standard deviation of 9.70 months. Appendix C contains a frequency distribution of the chronological ages for pupils in the sample.

Consideration of three other variables was also undertaken. Sixth, analysis of the marital status of parents of pupils revealed that 41/73 or 57.5 per cent of the pupils came from homes which have never been broken by death or divorce. Twenty-four or 32.9 per cent of the pupils come from homes which were broken by death or divorce. Seven or 9.6 per cent of the pupils came from homes which had been broken at one time but were, at the time of the study, intact because of remarriage to a step-parent. Seventh, the racial composition of the sample consisted of 62 per cent Negro and 38 per cent Caucasian.
Eighth, 90.4 per cent of the pupils in the sample came from Catholic families.

In summary, comparisons of this sample with the sample employed by W-K (1963, pp. 26-17) reveals their dissimilarity. First, the present sample is predominantly blue collar while the W-K sample was white collar. Second, the parents of the W-K sample were predominantly college graduates while the majority of the parents from the present study were high school graduates. Third, the families of pupils in the present sample were predominantly Protestant. Fourth, the area of residence of the W-K sample was suburban while that of the present sample was inner-city urban. Fifth, the present sample was largely Negro while the W-K sample was predominantly Caucasian.

Data Collection Procedure

The three examiners were introduced by the teacher of each class as persons interested in trying out various games which they had developed. One of the examiners explained to each class that participation in the games was not mandatory but the assistance of all members of the class would be very much appreciated. The pupils were also told that after they had completed all of the games they would be asked to take a test for the examiners. Considerable effort was made to assure the pupils that the scores from the test which they were to take would not go on their report card or in any way influence their
grades. In all interaction with the examinees the fact that the pupils were helping the examiners was emphasized.

It was necessary to present the W-K Test to the pupils as a new children's game in order to maintain a game-like atmosphere free from evaluation. For this reason, the W-K Test was administered before any of the more test-like procedures were introduced. Subjects were assigned to each of the three examiners by use of a table of random numbers. The W-K Test was individually administered by one of the three examiners in the manner prescribed by the battery authors. The standard directions developed by the authors were carefully employed in all cases.

The WISC was the second instrument administered for purposes of the present study. However, the two verbal learning tasks and one locus of control scale were individually administered to all members of the sample after completion of the creativity battery (see Asbury, 1970). The Similarities subtest was used to establish learning expectancy for the verbal learning tasks. Consequently, this subtest was administered in isolation before the rest of the WISC. With the exception of this subtest, all WISC subtests were administered in the order recommended by the test author.

The crayon and clay products were collected by the two respective classroom teachers. The teachers distributed the art materials and informed the pupils that they could make any type of product with the materials distributed to them.
The pupils were informed that the teacher would not help them because whatever they made should come from their own ideas rather than the teacher. All pupils were allowed to take as much time as they found necessary to complete each art product. Each of the three sets of artistic products was collected on different dates at least one week apart. The two sets of crayon drawings were collected before the clay products.

Description of Predictor Variables

The W-K Creativity Test was employed to secure a measure of creativity. This test was developed by W-K (1965) for use in their study. The battery consists of five subtests which are as follows: Instances, Alternative Uses, Similarities, Pattern Meanings, and Line Meanings. The test stimuli for Instances, Alternative Uses, and Similarities are verbal while the Line Meanings and Pattern Meanings are visual. All five subtests require verbal responses from the examinee which are recorded by the examiner during individual administration of the battery.

A pilot study revealed that administration of the full test battery required approximately two hours. Also, a large proportion of the pupils became fatigued during the last two subtests. An item analysis revealed that pupils made significantly fewer responses to items on the fifth subtest than to items on the second subtest. In this regard it should be noted that the order of the two subtests was counter-balanced
to control for test content. For example, when the Alternative Uses subtest was administered second pupils gave significantly \( p < .01 \) more responses than when this same test was administered fifth. This finding occurred despite the fact that a 10 minute rest break was allowed midway through the battery. Consequently, it was decided to shorten the battery by omission of several items. Shortening of the battery presented no particular problem because Wallach and Wing (1969) dropped one subtest and reduced the number of items on the other four subtests without adversely affecting any of the basic properties of the test.

Several items were dropped from the test to shorten it to the point where administration could be accomplished in one session of approximately 75 minutes. The items dropped were those with the least pull, i.e., those items from each subtest to which the 20 pupils in the pilot study made the fewest responses. All four items from the Instances subtest were retained. The items deleted on the remaining four subtests were as follows: Alternative Uses, 4, 6, and 8; Similarities, 7, 8, 9 and 10; Pattern Meanings, 3, 5, and 8; Line Meanings, 3, 4, and 8. The total number of items remaining on each of the five subtests after abbreviation of the battery was as follows: Instances, 4; Alternative Uses, 5; Similarities, 6; Pattern Meanings, 6; and Line Meanings, 6.

Because of the findings of Wallach and Wing (1969, p. 110) the battery was scored only for total number of asso-
ciations. The findings of this validation study suggested that number of associations is indicative of an individual's cognitive energy level or level of ideational activity.

The Wechsler Intelligence Scale for Children (Wechsler, 1949) was employed to secure a measure of general intelligence. In addition to a measure of general intelligence, the WISC also provides measures of verbal and performance ability. The following 11 subtests were administered to all subjects: Information, Comprehension, Arithmetic, Similarities, Vocabulary, Digit Span, Picture Completion, Picture Arrangement, Block Design, Object Assembly, and Coding.

Nine control variables were also included in the predictor battery. These were as follows: locus of control, chronological age, sex of pupil, intactness of home, race, religion, education of father, education of mother, and birth order.

The Bialer Locus of Control Scale (Bialer, 1961) was administered to all members of the sample in a group setting. The scale was scored so as to obtain an internal score. That is, high scores indicate an internal locus of control while low scores indicate an external locus of control.

Chronological age was determined from the date of birth recorded on each pupil's cumulative folder. The birth dates on the cumulative folder had been verified by examination of birth certificates by the school nurse. Chronological age to the nearest month was computed from the date of WISC admini-
stration. Since the administration of the WISC was accomplished during a nine day interval, the chances for systematic error from this source was minimized.

Data for the remaining seven variables was taken from the cumulative folders where it was usually available. However, in some cases it was necessary to contact parents directly to obtain the necessary information. This was procedure often necessary in the case of broken families. An arbitrary classification was employed for four of the discrete control variables. This procedure was as follows: sex (1 male, 0 female), home (1 intact home, 0 broken home), race (1 Caucasian, 0 Negro) and religion (1 Catholic, 0 non-Catholic). Education of mother and father was expressed in terms of years of formal education completed. Birth order was expressed in terms of the ordinal position of a given child's birth, i.e., first child, second child, etc. With regard to birth order, it should be noted that this variable is not necessarily an indication of family size.

Table 1 provides a summary of the 25 predictor variables as well as the notation used to describe each predictor variable in future references.

Description of Criterion Variables

Three sets of artistic products were completed by each child in the sample. Two sets of crayon drawings (Set I and Set II) and one set of clay products were collected in that respective order. Set I of the crayon drawings differed
<table>
<thead>
<tr>
<th>Predictor Variables by Instrument and/or Class</th>
<th>Notation for Predictor Variables</th>
</tr>
</thead>
<tbody>
<tr>
<td>Creativity Subtests:</td>
<td></td>
</tr>
<tr>
<td>Instances</td>
<td>X1</td>
</tr>
<tr>
<td>Alternative Uses</td>
<td>X2</td>
</tr>
<tr>
<td>Similarities</td>
<td>X3</td>
</tr>
<tr>
<td>Pattern Meanings</td>
<td>X4</td>
</tr>
<tr>
<td>Line Meanings</td>
<td>X5</td>
</tr>
<tr>
<td>WISC Subtests:</td>
<td></td>
</tr>
<tr>
<td>Verbal:</td>
<td></td>
</tr>
<tr>
<td>Information</td>
<td>X6</td>
</tr>
<tr>
<td>Comprehension</td>
<td>X7</td>
</tr>
<tr>
<td>Arithmetic</td>
<td>X8</td>
</tr>
<tr>
<td>Similarities</td>
<td>X9</td>
</tr>
<tr>
<td>Vocabulary</td>
<td>X10</td>
</tr>
<tr>
<td>Digit Span</td>
<td>X11</td>
</tr>
<tr>
<td>Performance:</td>
<td></td>
</tr>
<tr>
<td>Picture Completion</td>
<td>X12</td>
</tr>
<tr>
<td>Picture Arrangement</td>
<td>X13</td>
</tr>
<tr>
<td>Block Design</td>
<td>X14</td>
</tr>
<tr>
<td>Object Assembly</td>
<td>X15</td>
</tr>
<tr>
<td>Coding</td>
<td>X16</td>
</tr>
<tr>
<td>Bialer Locus of Control Scale:</td>
<td></td>
</tr>
<tr>
<td>Internal L-C</td>
<td>X17</td>
</tr>
<tr>
<td>Control Variables:</td>
<td></td>
</tr>
<tr>
<td>Chronological Age</td>
<td>X18</td>
</tr>
<tr>
<td>Sex</td>
<td>X19</td>
</tr>
<tr>
<td>Home</td>
<td>X20</td>
</tr>
<tr>
<td>Race</td>
<td>X21</td>
</tr>
<tr>
<td>Religion</td>
<td>X22</td>
</tr>
<tr>
<td>Father's Education</td>
<td>X23</td>
</tr>
<tr>
<td>Mother's Education</td>
<td>X24</td>
</tr>
<tr>
<td>Birth Order</td>
<td>X25</td>
</tr>
</tbody>
</table>
from Set II only in that the former was collected one week earlier.

The rating scale developed by Martinson and Seagoe (1967) provided the two criterion dimensions employed in the present study. These dimensions were originality and effectiveness of expression. However, the definitions of these respective dimensions were modified at the advice of the art consultant after a pilot study employing the Martinson-Seagoe definitions. The definitions employed in the present study were as follows: 1) **Originality**--originality, novelty and/or uniqueness in the ideas underlying the product; 2) **Effectiveness of Expression**--aesthetic quality, exquisite quality in community, and/or an impression of beauty in the use of elements or media. The extent to which the original definitions of Martinson and Seagoe were modified for purposes of the present study can be readily determined by consulting their monograph (pp. 11-46).

In accordance with the Martinson-Seagoe procedure, a nine category rating scale was employed. However, the category descriptions were changed so that the scale ranged from 9 (superior) through 5 (average) to 1 (inferior). The nine categories were as follows: superior, very good, good, high average, average, low average, poor, very poor, and inferior.

The present study further differs from that of Martinson and Seagoe in that the judges made independent ra-
tings for the Originality and Effectiveness of Expression dimensions. The ratings for these two dimensions were not combined as in the Martinson-Seagoe study. Rather, the two dimensions were treated separately in accordance with the analytic orientation of Underwood (1957).

Furthermore, the pupils remained totally anonymous to the judges in the present study as they were known only through a system of identification numbers which was changed from one set of products to another. In the Martinson-Seagoe study pupil names were used rather than identification numbers. Use of names was not a critical consideration in the Martinson-Seagoe procedure since different judges rated each set of products. However, when the same judges rate more than one set of products by the same group of pupils the use of staggered identification numbers is essential.

Finally, the actual rating procedure employed in the present study differed from that of Martinson and Seagoe in that a stacking procedure was employed. That is, the judges were instructed to arrange the products into nine groups corresponding to the nine categories of the rating scale. According to the classification system of Torgerson (1958) the rating procedure employed in the present study can best be described as a subjective estimate method. In terms of Torgerson's (1958, pp. 66-68) classification system the specific data collection procedure can be described as a multiple-stimuli method with limited categories.
However, it should be clearly understood that the rating procedure employed in this study is related to Torgerson's classification system only for purposes of illustration. The rating procedure employed in the present study cannot technically be described as a subjective estimate method as the judges were instructed to assume a particular distribution of stimuli (Torgerson, 1958, p. 67). Since a detailed consideration of scaling is beyond the scope of the present study, the Martinson-Seagoe procedure of instructing the judges to place an equal number of products in each category was followed. Thus, the reader should be aware of the implications inherent in assuming a rectangular distribution of criterion ratings and interpret the findings accordingly.

Four judges recommended by a faculty consultant from the Department of Art completed ratings of the three sets of artistic products produced by each pupil. Each judge was paid the sum of twenty-five dollars ($25.00) for rating the three sets of products.

The three sets of products were presented to the judges according to a uniform procedure. The four judges worked independently in the three rooms containing the products to be rated. All communication between judges was eliminated during the rating procedure. The judges were, however, allowed to chat with the art consultant and investigator during periods of time when they were waiting to be conducted to the next set of products.
Originality ratings were completed for each set of products in a fixed order. After each judge completed his ratings of the Originality dimension for all three sets of products, he then rated the three sets of products on the Effectiveness of Expression dimension.

Appendix D contains a detailed description of the rating procedure employed in the study.

Table 2 summarizes the criterion variables and provides a notation system used for indicating the criterion variables throughout the remainder of the dissertation.
Table 2: Notation for Criterion Variables Employed in the Study

<table>
<thead>
<tr>
<th>Criterion Variable</th>
<th>Notation</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Crayon Drawings - Set I</strong></td>
<td></td>
</tr>
<tr>
<td>Originality Ratings</td>
<td>C1</td>
</tr>
<tr>
<td>Effectiveness of Expression Ratings</td>
<td>C5</td>
</tr>
<tr>
<td><strong>Crayon Drawings - Set II</strong></td>
<td></td>
</tr>
<tr>
<td>Originality Ratings</td>
<td>C2</td>
</tr>
<tr>
<td>Effectiveness of Expression Ratings</td>
<td>C6</td>
</tr>
<tr>
<td><strong>Clay Products</strong></td>
<td></td>
</tr>
<tr>
<td>Originality Ratings</td>
<td>C3</td>
</tr>
<tr>
<td>Effectiveness of Expression Ratings</td>
<td>C4</td>
</tr>
</tbody>
</table>
CHAPTER V
RESULTS AND DISCUSSION

This chapter is devoted to examination of the results of the study in relationship to the four research questions formulated for purposes of the study. For purposes of clarity in exposition the chapter is divided into five major sections. The major sections were divided into subsections as necessary. The first section was devoted to examination of data preliminary to the consideration of the major research questions. The four remaining sections were devoted to examination of the respective research questions.

The data analyses employed to obtain the results reported in each major section of the chapter are described in the introductory paragraphs of each respective section or subsection. In this respect it should be noted that the analyses employed were those which, in the opinion of the writer, were most appropriate. However, Appendix E contains a summary of all data used to obtain the findings presented in this chapter so that researchers preferring other statistical techniques may find the necessary data readily available.
Preliminary Data

This section is devoted to consideration of two major trends in the data not directly relevant to the research questions which were formulated a priori. That is, the findings discussed in this section became evident at diverse points in the sequence of data analyses specified for answering the research questions. Inspection of the raw score frequency distributions for the five subtests of the W-K Test suggested the need to focus attention on the form of the standard score distributions. In specific, the respective means, standard deviations, and indices of skewedness were computed for each subtest.

The second topic of the present section became evident in much the same way. That is, examination of the intercorrelation matrix for the predictor and criterion variables revealed the existence of a strong negative relationship between birth order and each of the five creativity subtests. The focus of the data analyses was such that this important relationship might easily be overlooked. In fact, the very nature of the relationship noted above minimized the chances that the birth order variable would be included in the regressions to be discussed later.

Thus, the findings discussed in this section bear an important overall relationship to the research questions to
be discussed in the remainder of the chapter. In addition, the nature of the findings is such that they should be considered before discussion of the findings directly relevant to the respective research questions.

**Distribution of Scores on the Creativity Subtests**

The median, mean, and standard deviation were computed for the scores on each creativity subtest. These values were then used to compute an index of skewedness in the manner described by Garrett (1958, pp. 99-101). The formula for this index is as follows:

\[ SK = \frac{3(X - \text{Mdn.})}{S} \]

This index provides a rough indication of the magnitude of skewedness as well as its direction. Table 3 shows the number of items in each subtest as well as the median, mean, standard deviation, and skewedness index.

Examination of the data in Table 3 clearly reveals that the distribution of scores on each of the five creativity subtests exhibited substantial positive skewedness. The mean of each subtest was larger than the median. More specifically, the values of the skewedness were all positive and ranged from moderate to high. The degree of positive skewedness was greater for the two subtests consisting of visual stimuli than for the three subtests in which verbal stimuli were employed. Examination
Table 3: Descriptive Statistics for the Five Creativity Subtests

<table>
<thead>
<tr>
<th>Subtest</th>
<th>Number of Items in Subtest</th>
<th>Mdn.</th>
<th></th>
<th>S</th>
<th>SK</th>
</tr>
</thead>
<tbody>
<tr>
<td>Instances</td>
<td>4</td>
<td>48.25</td>
<td>55.14</td>
<td>31.76</td>
<td>0.65</td>
</tr>
<tr>
<td>Alternative Uses</td>
<td>5</td>
<td>29.27</td>
<td>32.04</td>
<td>15.68</td>
<td>0.53</td>
</tr>
<tr>
<td>Similarities</td>
<td>6</td>
<td>28.66</td>
<td>31.03</td>
<td>15.17</td>
<td>0.47</td>
</tr>
<tr>
<td>Pattern Meanings</td>
<td>6</td>
<td>23.76</td>
<td>28.84</td>
<td>16.56</td>
<td>0.92</td>
</tr>
<tr>
<td>Line Meanings</td>
<td>6</td>
<td>25.43</td>
<td>33.26</td>
<td>22.19</td>
<td>1.06</td>
</tr>
</tbody>
</table>
of Appendix F, which contains z-score distributions for the five creativity subtests supports the interpretations drawn above.

These data indicate that the majority of the pupils in the sample gave relatively few associations while a few pupils gave a much larger number of associations. The fact that the shape of the score distributions for the five subtests were similar suggests that creativity, as defined by the W-K operations, is not normally distributed. The similarity of the distributions for the five subtests might also be cited to support the dimensionality of the W-K operations.

**Birth Order and Performance on the W-K Test**

The relationship between the subtest of the W-K Test and birth order was investigated through examination of the intercorrelations between the birth order variable and scores on the five respective creativity subtests. The data discussed in the present subsection were obtained from the predictor-criterion intercorrelations matrix provided by the WHEWH Program (Wherry and Wherry, 1968) designed for use on the IBM 7094. Appendix G contains the intercorrelation matrix for all predictor and criterion variables employed in the study.
Examination of the correlations between birth order and scores on the respective creativity subtests revealed a strong negative relation in each case. The correlations between birth order and the various creativity subtests were as follows: Instances, -0.62; Alternative Uses, -0.58; Similarities, -0.55; Pattern Meanings, -0.60; and Line Meanings, -0.49. All five correlation coefficients attained significance at beyond the 0.01 level. This finding indicates that being an older child within a given family and/or being from a small family is an important correlate of generating a large number of associations to both verbal and visual stimuli.

The findings of the present subsection stand in uncertain relationship to the majority of the literature. Several factors account for this state of affairs. First, the birth order variable has not been previously related to the W-K creativity operations. In the absence of research employing similar operations one is considerably handicapped by the lack of consensual validation. Second, there is a dearth of research concerning the relationship between the W-K operations and other operations purported to measure creative potential, e.g., well known divergent production tests such as those of Guilford and Torrance. Third, researchers employing divergent production tests with children have usually neglected to relate their operations to biographical variables such as
birth order. Given this state of affairs, it makes little sense to attempt to summarize the relationship between birth order and other creativity tests.

However, the situation is considerably more promising when one concentrates on criterion based studies. Cattell and Brimhall (1921) examined selected family history variables for several hundred eminent scientists and found a high proportion of first-born children. Roe (1953) also reported a high proportion of first-borns in her intensive study of a small sample of eminent scientists.

The result of MacKinnon's (1960) study of eminent architects stood in sharp contrast to those studies noted above. MacKinnon observed that his highly creative architects tended to come from larger than average families and to report warm friendly relations with siblings. Finally, Datta (1967) reported a study of the correlates of scientific attainment among male high school students. The results of her study failed to attribute special significance to either family size or birth order. Thus, it is evident that the importance of birth order is highly equivocal even when one confines attention to criterion based studies.

Some individuals have expressed the opinion that the predominance of first-borns among eminent persons is, in reality, an artifact of inappropriate control procedures. Among those advocating this interpretation are Schachter (1963) and Christie
(1970). Schachter's interpretation was suggested by the predominance of descriptive studies of eminent individuals across various fields. Christie's position was inspired by similar considerations as well as the studies of Datta and MacKinnon noted above.

The findings on the present study sharply contradict the interpretations of Schachter and Christie. Yet, it is obvious that the issue under consideration can be resolved only by a wide range of criterion-based studies.

Question 1: To what extent will the discriminant validity from general intelligence and high subtest intercorrelations reported by W-K and others be generalized to pupils in the present sample with a different social, ethical, and economic background?

The present section is devoted to answering the first research question posed for purposes of the study. Intercorrelations between and among the intelligence and creativity subtests were obtained from the WHEWH program (Wherry and Wherry, 1968). Appendix G contains the full predictor-criterion intercorrelation matrix with values of \( r \) rounded to two decimals.

The data summarized in Table 4 shows a strong positive relationship among the creativity subtests. Nine of the ten intercorrelations attained significance at beyond the 0.01 level and the remaining one attained significance
Table 4: Intercorrelations Among the Five Creativity Subtests

<table>
<thead>
<tr>
<th>Creativity Subtests:</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Instances</td>
<td>1.000</td>
<td>.767</td>
<td>.673</td>
<td>.548</td>
<td>.467</td>
</tr>
<tr>
<td>2. Alternative Uses</td>
<td>1.000</td>
<td>.768</td>
<td>.563</td>
<td>.400</td>
<td></td>
</tr>
<tr>
<td>3. Similarities</td>
<td>1.000</td>
<td>.632</td>
<td>.627</td>
<td></td>
<td></td>
</tr>
<tr>
<td>4. Pattern Meanings</td>
<td>1.000</td>
<td>.830</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5. Line Meanings</td>
<td>1.000</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Note: For 71 df and 10 sequential t-tests, values of 0.354 and 0.413 are significant at the 5% and 1% levels.
at beyond the 0.05 level. The average intercorrelation among the ten subtests obtained by averaging the respective values after an r-to-z transformation and converting the average z-value back to r was 0.623. Squaring the values in Table 4 clearly indicates that the five creativity subtests share substantial common variance. Thus, W-K's assertion that their operations define a unitary dimension of cognitive ability seems justified.

The relationship between the creativity and intelligence subtests was also investigated through an analysis of subtest intercorrelations. Table 5 contains the intercorrelations between the five creativity and eleven intelligence subtests.

Examination of data in Table 5 reveals the overall divergence of scores on the creativity subtests from general intelligence as measured by WISC subtests. The average intercorrelation among the creativity and intelligence subtests obtained after averaging the respective z-values and converting the average z-value back to r was 0.08. Since maintenance of an adequate protection level for fifty sequential t-tests is unrealistic, examination of the intercorrelation matrix for patterns were found. Both the Object Assembly and Coding subtests showed slight to moderate positive correlations with all five creativity subtests.
<table>
<thead>
<tr>
<th>Creativity Subtest</th>
<th>I</th>
<th>C</th>
<th>A</th>
<th>S</th>
<th>V</th>
<th>DS</th>
<th>PC</th>
<th>PA</th>
<th>BD</th>
<th>OA</th>
<th>C</th>
</tr>
</thead>
<tbody>
<tr>
<td>Instances</td>
<td>.017</td>
<td>-.124</td>
<td>.062</td>
<td>.006</td>
<td>-.011</td>
<td>-.058</td>
<td>.086</td>
<td>-.202</td>
<td>-.018</td>
<td>.277</td>
<td>.153</td>
</tr>
<tr>
<td>Alternative Uses</td>
<td>.018</td>
<td>.030</td>
<td>.011</td>
<td>-.066</td>
<td>.037</td>
<td>.086</td>
<td>.134</td>
<td>.048</td>
<td>-.036</td>
<td>.348</td>
<td>.113</td>
</tr>
<tr>
<td>Similarities</td>
<td>.073</td>
<td>.116</td>
<td>.147</td>
<td>.082</td>
<td>.192</td>
<td>.198</td>
<td>.158</td>
<td>.098</td>
<td>.115</td>
<td>.389</td>
<td>.287</td>
</tr>
<tr>
<td>Pattern Meanings</td>
<td>-.023</td>
<td>-.057</td>
<td>.017</td>
<td>-.073</td>
<td>.064</td>
<td>-.015</td>
<td>.039</td>
<td>.072</td>
<td>.004</td>
<td>.269</td>
<td>.204</td>
</tr>
<tr>
<td>Line Meanings</td>
<td>-.029</td>
<td>.002</td>
<td>.027</td>
<td>.040</td>
<td>.119</td>
<td>-.018</td>
<td>.051</td>
<td>.033</td>
<td>-.014</td>
<td>.188</td>
<td>.272</td>
</tr>
</tbody>
</table>
These data support W-K's assertion that their creativity operations measure a dimension of cognitive development with discriminant validity from general intelligence. Positive relationship observed between the Object Assembly and Coding subtests and the five creativity subtests will become clearer in the discussion of the factorial composition of the predictor-criterion matrix.

Table 6 contains the intercorrelations among the eleven intelligence subtests. The average intercoorelation among these subtests obtained through an appropriate r-to-z transformation was 0.239. From these data it can be seen that the creativity subtests show a substantially higher average intercorrelation than the intelligence subtests.

In summary, the first research question can be positively answered without equivocation. That is, the discriminant validity from general intelligence and dimensionality obtained by W-K (1965), J. Ward (1967), W. Ward (1966, 1968), Pankove (1967), Cropley (1968), Fee (1968) and Wallach and Wing (1969) generalized to the present sample even though it was composed of children with different social, economic, and ethnic characteristics than those of the studies cited above. Alternatively stated, the findings discussed in this section have the effect of providing further evidence that the W-K creativity operations define a basic behavioral phenomena which is to be found across a wide range of samples ranging
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>1.000</td>
<td>.419</td>
<td>.448</td>
<td>.311</td>
<td>.571</td>
<td>.220</td>
<td>.261</td>
<td>.306</td>
<td>.175</td>
<td>.197</td>
<td>.118</td>
<td>.118</td>
</tr>
</tbody>
</table>
from the early primary grades through early adulthood in age.

Question 2: Will the four judges achieve a significant degree of concordance in their ratings of the artistic products submitted to them for judgement?

The data discussed in this subsection were analyzed by use of the CHIRAT Program (Wherry, 1970) designed for use with the IBM 360. This program provides a coefficient of concordance \(W\), a \(x^2\)-test for the significance of the obtained \(W\), an average-rho \(\bar{r}_s\), and an average-rho corrected for number of judges by the Spearman-Brown formula \(\bar{r}_{snn}\). These statistics were employed to determine the extent of interjudge agreement on each of the six creativity criteria.

An alternative to the present strategy for data analysis consists of computing Pearson \(r\)'s between the ratings of each pair of judges, averaging these \(r\)'s, and then correcting the average \(r\) by the Spearman-Brown formula. A detailed discussion of the relative merits of parametric versus nonparametric statistics is beyond the scope of the present study. However, it should suffice to note that the strategy employed is conservative in that one obtains lower estimates of relationship by use of nonparametric methods. Second, power efficiency was not a major concern as there was no interest in detecting weak relationships. It was necessary that the judges' ratings show a high degree of concordance before summation of their
separate ratings to obtain a composite criterion rating
could be justified.

Examination of the statistics reported in Table
7 clearly shows that a highly significant degree of concord-
ance did in fact exist among the ratings of the four judges.
This was true for all six creativity criteria. The coeffi-
cients of concordance (W's) for each criterion attained
significance at beyond the 0.001 level as indicated by their
respective $x^2$-tests. These data reveal that the four judges
achieved a highly significant degree of agreement.

The values of average-rho ($\bar{r}_S$) may be described as
the average of all possible combinations of rank-order
correlations between the four judges' ratings of a given
criterion. However, since the rankings are considered only
two at a time $\bar{r}_S$ does not indicate the extent to which the
ratings of the four judges agree when considered as a unit.
The Spearman-Brown prophecy formula was applied to the values
of $\bar{r}_S$ to determine the reliability of the combined ratings of
the four judges.

Examination of the statistical indices of inter-
judge agreement contained in Table 7 suggests that summing
the ratings of the four judges to obtain a composite rating
for each criterion was justified. On both criterion dimensions
interjudge agreement was higher for clay products than for
crayon drawings. Also, both the first and second set of
<table>
<thead>
<tr>
<th></th>
<th>W</th>
<th>$x^2$</th>
<th>df</th>
<th>p</th>
<th>$\bar{r}_s$</th>
<th>$\bar{r}_{s4IV}$</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Crayon Drawings - Set I</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Originality</td>
<td>0.601</td>
<td>173.06</td>
<td>72</td>
<td>.001</td>
<td>0.468</td>
<td>0.799</td>
</tr>
<tr>
<td>Effectiveness of</td>
<td>0.463</td>
<td>133.40</td>
<td>72</td>
<td>.001</td>
<td>0.284</td>
<td>0.614</td>
</tr>
<tr>
<td>Expression</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Crayon Drawings - Set II</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Originality</td>
<td>0.600</td>
<td>172.67</td>
<td>72</td>
<td>.001</td>
<td>0.466</td>
<td>0.777</td>
</tr>
<tr>
<td>Effectiveness of</td>
<td>0.539</td>
<td>155.19</td>
<td>72</td>
<td>.001</td>
<td>0.385</td>
<td>0.715</td>
</tr>
<tr>
<td>Expression</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Clay Products</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Originality</td>
<td>0.695</td>
<td>200.26</td>
<td>72</td>
<td>.001</td>
<td>0.594</td>
<td>0.854</td>
</tr>
<tr>
<td>Effectiveness of</td>
<td>0.718</td>
<td>206.67</td>
<td>72</td>
<td>.001</td>
<td>0.624</td>
<td>0.869</td>
</tr>
</tbody>
</table>
crayon drawings the judges' ratings of the Originality dimension agreed better than their ratings of the Effectiveness of Expression dimension.

These findings suggest that the judges were able to rate the Originality dimension of crayon drawings more accurately than the Effectiveness of Expression dimension. This hypothesis is supported by the comments made by two of the judges. Two judges volunteered comments to the effect that Originality was much easier and more pleasant to judge than Effectiveness of Expression.

The original design of the study specified the collection of two sets of crayon products so that these could be summed to obtain a more reliable criterion measure. This procedure was deemed advisable since Martinson and Seagoe (1967) found that interjudge agreement was considerably lower for drawings than for clay products. In specific, it was planned to sum the Originality ratings for Sets I and II of crayon drawings to obtain a more reliable criterion measure. A similar procedure was also planned for Effectiveness of Expression ratings of crayon drawings. The data analysis next described in this subsection was conducted to determine whether summation of the Originality and Effectiveness of Expression ratings for the two sets of crayon drawings could be justified.

Data analysis revealed that one could not justify
summing the judges' ratings for the two sets of crayon drawings. This was true for both the Originality and Effectiveness of Expression dimensions. Summation of the ratings could be justified only if, on a given dimension, the ratings for Set I correlated highly with the ratings for Set II. This did not occur. The product-moment correlation between Originality ratings for Sets I and II of crayon drawing was only 0.146. This value neither attained statistical significance nor indicated the existence of substantial common variance in the two criteria. The correlation between Effectiveness of Expression ratings for Sets I and II of crayon was only 0.200. This value did not attain statistical significance. Furthermore, it indicated minimal variance common to the two sets of ratings. Given these findings, summation of the ratings for the two sets of crayon drawings could not be justified.

Consideration of these findings in relation to the research question of concern in this section reveals a baffling state of affairs. The research question posed for purposes of the study was clearly answered affirmatively since the judges achieved a highly significant degree of concordance for each of the six sets of artistic products. Yet, a subsidiary analysis of the correlation between ratings of the two sets of crayon drawings for both dimensions revealed the presence of a problem in the procedure employed to develop the criteria. While the judges were able to agree
well on their ratings for each set of crayon drawings, their ratings of a given dimension for the two sets of crayon drawings did not agree.

In terms of traditional psychometric theory this situation is analogous to a test with high internal consistency reliability but extremely poor test-retest reliability. With the data at hand one is unable to determine if the standards of the judges changed or if the second set of products produced by the children actually differed from the first. It is, of course, also possible that both of these factors interacted to produce the situation at hand.

However, the fact that all four judges agreed in their ratings of both sets of products on both dimension suggests that the product may have varied systematically. In other words, the first set of product procured by the children may well have differed from the second set. This hypothesis is consistent with recent advances in psychometric theory. For example, Cattell and Butcher (1968) note that some traits fluctuate so much that "...a test may be extra ordinarily valid for a particular trait at a particular moment but deficient in long-term prediction of this trait" (p. 98).

In any case, the inconsistency noted above serves only to further emphasize the need for additional research in the area of criterion development. This is true because the issue raised above, like most issues in creativity research, can be resolved only through careful research. To answer this
question it will be necessary to collect samples of the crayon
drawings of children across various time intervals and submit
these drawings to a panel of judges for evaluation on various
dimensions.

Question 3: What is the factorial
composition of the
predictor and criterion
variables employed in the
study?

The Wherry Hierarchical Solution

The WHEWH Program (Wherry and Wherry, 1968) designed
for use with the IBM 7094 was applied to the predictor-criterion
intercorrelation matrix to obtain the results presented in this
section. This program consists of Wherry's most recent computa-
tional procedure for obtaining the hierarchical factor solu-
tion which he originally published in Psychometrika (Wherry,
1959).

The initial step in the Wherry hierarchical solu-
tion consists of obtaining cluster sums and correlations
by the multiple group factor method. This method is applied
sequentially to the intercorrelations among clusters until
the final cluster matrix consists of one single cluster.
Wherry assumes that if all the common variance is removed
from the clusters through sequential application of the
multiple group factor method then the clusters should demon-
strate simple structure with respect to each other.

The final cluster obtained should, then represent
a general factor with loadings on the original variables. At
the same time, the sub-general factors should exhibit simple structure with respect to one another and have loadings on all variables in the lower order clusters composing them. Each cluster is assumed to possess unit variance in the computation of specific loadings while communalities are employed in the computations involved in multiple cluster factorization.

At this stage the analysis provides the highest level unitary factor cluster matrix with a set of factor loadings for each variable and specific factor loadings for the respective lower order clusters represented in the matrix. In addition, cluster sums for remaining lower order clusters are available as well as the multipliers determined by taking the reciprocal of the square root of the sum of the cluster sums for the variables in each respective cluster. The cluster sums are those originally used to obtain the cluster correlation matrix at that level.

The second stage of Wherry's method consists of a back-solution phase designed to complete the factor pattern for the sub-general and group factors. Wherry (1959, p. 46) provides a succinct summary of the remaining steps:

1. Using the highest order cluster correlation matrix with ones in the diagonals and the general and specific factor loadings as criteria, obtain the beta weights which would predict these factors from the clusters.
2. These beta weights are then multiplied by the multipliers referred to above, and these products are used as a transformation matrix.
3. These transformation weights are used to multiply the cluster sums of the previous analysis, which yields general and sub-general
factor loadings for each of the next lower order clusters.

(4) The communality is computed from these factor loadings for each cluster at this new level, subtracted from unity, and the square root is taken to obtain a specific factor loading for the cluster.

(5) Steps (1) through (4) are repeated for each successive lower order matrix until the original correlation matrix for the variables is involved.

The advantages of the Wherry hierarchial solution are obvious. First, it completely eliminates the need for rotation to oblique simple structure at each stage of the analysis. Second, it eliminates dealing with inferred higher order factors necessary in oblique solutions. Third, it maintains orthogonal factors at all levels of the hierarchy. Fourth, it allows for the projections of all variables on factors at all levels in the hierarchy so that the overall factor pattern is available.

Other psychologists have advocated a hierarchial theory of intelligence. However, Wherry's (1959) approach is unique in that it provides both a sound conceptual basis and a convenient computational procedure for employment in a research setting. Little progress could be made in testing and elaborating hierarchial theories of intelligence until the Wherry hierarchial solution became available. In fact, the Wherry hierarchial solution provides an excellent vehicle for testing hypotheses concerning the structure-of-intellect. If, as some theoreticians maintain, intelligence is composed of orthogonal primary factors then a factor
structure congruent with such a hypothesis should be obtained from a hierarchial solution. In specific, if the variables subjected to analysis do not cluster in a hierarchial arrangement then a factor structure congruent with a hypothesis of independent primary factors would be obtained.

Perhaps the most intriguing implication of Wherry's hierarchial paradigm consists of the challenge to develop and test hypotheses concerning the growth and development of cognitive functioning. For example, the hierarchial solution presents an ideal tool for testing the differentiation hypothesis of cognitive ability. Hierarchial factor analysis of abilities at various ages should present a much clearer picture of the pattern for the growth and decline of intelligence than is now available. Furthermore, the application of hierarchial analysis in conjunction with ANOVA techniques in experimental studies of cognition should serve to clarify the nature of any charges brought about by the treatment conditions.

The results of the data analysis presented in this section is presented in the following manner. Each factor is first discussed individually before a synthesis of the overall structure of abilities for the sample is approached. The tests defining each respective factor are listed and discussed under an appropriate caption for that factor. Table 8 contains the hierarchial factor matrix obtained from application of the Wherry hierarchial solution. As such, this table constitutes
Table 8: Hierarchical Factor Matrix

<table>
<thead>
<tr>
<th>Variable</th>
<th>Higher Order Factors</th>
<th>Primary Factors</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>g</td>
<td>c</td>
</tr>
<tr>
<td>X1</td>
<td>.06</td>
<td>.29</td>
</tr>
<tr>
<td>X2</td>
<td>.19</td>
<td>.24</td>
</tr>
<tr>
<td>X3</td>
<td>.29</td>
<td>.39</td>
</tr>
<tr>
<td>X4</td>
<td>.05</td>
<td>.41</td>
</tr>
<tr>
<td>X5</td>
<td>.01</td>
<td>.44</td>
</tr>
<tr>
<td>X6</td>
<td>.63</td>
<td>.00</td>
</tr>
<tr>
<td>X7</td>
<td>.59</td>
<td>.01</td>
</tr>
<tr>
<td>X8</td>
<td>.51</td>
<td>.18</td>
</tr>
<tr>
<td>X9</td>
<td>.27</td>
<td>.11</td>
</tr>
<tr>
<td>X10</td>
<td>.67</td>
<td>.17</td>
</tr>
<tr>
<td>X11</td>
<td>.27</td>
<td>.07</td>
</tr>
<tr>
<td>X12</td>
<td>.38</td>
<td>.16</td>
</tr>
<tr>
<td>X13</td>
<td>.35</td>
<td>.00</td>
</tr>
<tr>
<td>X14</td>
<td>.34</td>
<td>.23</td>
</tr>
<tr>
<td>X15</td>
<td>.45</td>
<td>.21</td>
</tr>
<tr>
<td>X16</td>
<td>.14</td>
<td>.16</td>
</tr>
<tr>
<td>X17</td>
<td>.25</td>
<td>-.11</td>
</tr>
<tr>
<td>X18</td>
<td>-.35</td>
<td>-.06</td>
</tr>
<tr>
<td>X19</td>
<td>.03</td>
<td>-.02</td>
</tr>
<tr>
<td>X20</td>
<td>-.03</td>
<td>-.15</td>
</tr>
<tr>
<td>X21</td>
<td>.00</td>
<td>.12</td>
</tr>
<tr>
<td>X22</td>
<td>.27</td>
<td>.05</td>
</tr>
<tr>
<td>X23</td>
<td>.08</td>
<td>.10</td>
</tr>
<tr>
<td>X24</td>
<td>.08</td>
<td>.05</td>
</tr>
<tr>
<td>X25</td>
<td>-.06</td>
<td>-.29</td>
</tr>
<tr>
<td>C1</td>
<td>-.02</td>
<td>.20</td>
</tr>
<tr>
<td>C2</td>
<td>-.04</td>
<td>.29</td>
</tr>
<tr>
<td>C3</td>
<td>-.23</td>
<td>.44</td>
</tr>
<tr>
<td>C4</td>
<td>-.23</td>
<td>.49</td>
</tr>
<tr>
<td>C5</td>
<td>.05</td>
<td>.34</td>
</tr>
<tr>
<td>C6</td>
<td>.16</td>
<td>.29</td>
</tr>
</tbody>
</table>
the source of the factor coefficients for the variables loading on each respective factor discussed in this section. The following subsection is devoted to a description and interpretation of the factors defined by the hierarchical solution.

Description and Interpretation of Factors

Examination of the variables loading Factor g clearly indicate that it is equivalent to what has traditionally been known as general intelligence. The variables with the highest positive loadings are all WISC subtests. More important, however, is the fact that general intelligence has long been operationally defined as the overlap among diverse assessors of intelligence. The configuration of variables loading Factor g is clearly congruent with this conventional concept of general intelligence. Finally, the pattern of variables loading Factor g as well as the relative magnitude of the respective loadings closely parallels those reported by other investigators, e.g., Cohen (1959).

Factor g is, then, clearly defined by a configuration of positive loadings on the part of WISC subtests. The WISC subtests loading highest on the g-factor were Vocabulary, Information, Comprehension, and Arithmetic. Again, the results are in substantial agreement with Cohen's (1959) oblique analysis in which WISC subtest scores were correlated with scores on a second-order general factor composed of a linear combination of WISC subtest scores. Thus, the findings of the present study have the effect of supporting Cohen's con-
### Table 8.1: Factor g - General Intelligence

<table>
<thead>
<tr>
<th>X3:</th>
<th>Similarities (W-K)</th>
<th>.29</th>
</tr>
</thead>
<tbody>
<tr>
<td>X6:</td>
<td>Information</td>
<td>.63</td>
</tr>
<tr>
<td>X7:</td>
<td>Comprehension</td>
<td>.59</td>
</tr>
<tr>
<td>X8:</td>
<td>Arithmetic</td>
<td>.51</td>
</tr>
<tr>
<td>X9:</td>
<td>Similarities (WISC)</td>
<td>.27</td>
</tr>
<tr>
<td>X10:</td>
<td>Vocabulary</td>
<td>.67</td>
</tr>
<tr>
<td>X11:</td>
<td>Digit Span</td>
<td>.27</td>
</tr>
<tr>
<td>X12:</td>
<td>Picture Completion</td>
<td>.38</td>
</tr>
<tr>
<td>X13:</td>
<td>Picture Arrangement</td>
<td>.35</td>
</tr>
<tr>
<td>X14:</td>
<td>Block Design</td>
<td>.34</td>
</tr>
<tr>
<td>X15:</td>
<td>Object Assembly</td>
<td>.45</td>
</tr>
<tr>
<td>X16:</td>
<td>Coding</td>
<td>.14</td>
</tr>
<tr>
<td>X17:</td>
<td>Locus of Control (internal)</td>
<td>.25</td>
</tr>
<tr>
<td>X18:</td>
<td>Chronological Age</td>
<td>-.35</td>
</tr>
<tr>
<td>X22:</td>
<td>Religion</td>
<td>.27</td>
</tr>
<tr>
<td>C3:</td>
<td>Originality (C.P.)</td>
<td>-.23</td>
</tr>
<tr>
<td>C4:</td>
<td>Effect of Expression (C.P.)</td>
<td>-.23</td>
</tr>
</tbody>
</table>
clusion that, "...the 'essentially verbal' tests, particularly Vocabulary and Information, are consistently the best measures of G..." (1959, p. 289).

The remaining WISC subtests showed relatively lower loadings on the g-factor. Subtests showing moderate loadings on this factor were Object Assembly, Picture Completion, Picture Arrangement, and Block Design. The Similarities and Digit Span subtests showed slight positive loadings while Coding showed only a negligible loading.

As indicated above, the results of the present study agree with those of Cohen's (1959) study by indicating that the Vocabulary, Information, Comprehension, and Arithmetic subtests constitute the best measure of g. However, comparison of the absolute magnitude of subtest loadings could not be meaningfully achieved since the results of Cohen's analysis attribute a much higher proportion of subtest variance to the g-factor than the result of the present hierarchial solution would justify. This discrepancy is true of most subtests concerned.

Cohen probably overestimated the proportion of variance accounted for by his second-order g-factor as he computed correlations between the respective subtest scores and g-factor scores which consisted of a weighted linear combination of WISC subtest scores.

The weakness in Cohen's study stemmed from his testing factor scores are factorially pure. That is, the coeffi-
cients for each variable included in his equation for g-factor scores serve as multipliers for all the factorial components of that respective subtest. This procedure necessarily leads to contamination of g-factor scores with elements of the primary factors. Elements of the primary factors are included in scores for both the g-factor and the respective subtests. Therefore, squaring the correlations between scores on the g-factor and the respective subtests overestimate the proportion of subtest variance accounted by g.

Thus, the results of the present study suggest that the method of data analysis employed by Cohen (1959) led to an overestimation of the proportion of subtest variance explained by the g-factor. The investigator would take the position that the squared loadings of the respective WISC subtests on the g-factor provide the best estimate of the proportion of variance accounted for by general intelligence.

The g-factor was loaded by several variables other than the WISC subtests by which it was defined. The negative loading by chronological age (C.A.) is particularly noteworthy. Inspection of the data in Appendix E suggests the negative loading by C.A. on the g-factor reflects a substantial tendency for overage-in-grade pupils to score lower on verbal subtests of the WISC. This interpretation agreement with Cohen's (1959) study which suggests that general intelligence does not decrease with age.
This finding suggests that caution is necessary in interpreting correlations between C.A. and other variables. That is, C.A. cannot be interpreted as a control for maturation as originally intended. Rather, the relationship between C.A. and other variables must be regarded as some undetermined interaction between the contamination effects induced by retaining the overage-in-grade pupils and normal developmental changes.

However, the overage-in-grade pupils were retained in the sample despite the difficulty noted above. These pupils were retained so that the regression equations for the creativity criteria might reflect a realistic state of affairs frequently found in both public and parochial schools. That is, many schools maintain a policy of retaining pupils not achieving at a level commensurate with their age mates. For this reason a regression equation based on the full range of pupils should stand a better chance of holding up in cross-validation studies employing heterogeneous samples.

Investigators desiring information of a different nature may readily perform analyses of their own choice using the data in Appendix E.

The Similarities subtest of the W-K also loaded the g-factor positively. The magnitude of this loading was such that the g-factor accounts for about nine percent of the variance in the Similarities subtest. This loading is not difficult to understand since similarities items have a long history
in intelligence testing. A threshold effect is suggested as it appears that a child must possess some minimal degree of general intelligence before he is able to generate a plentiful supply of similarities to the W-K items employed in the study. From a broader perspective it seems reasonable to surmise that the general intelligence threshold varies with the degree of abstraction required in determining the essential similarities between the stimulus words.

The threshold hypothesis may well extend to tasks other than similarities as W. Ward (1966, pp. 103-114) observed a comparable effect with his Pictures test which was scored for number of unique and non-unique associations.

Another variable loading the g-factor positively was the internal locus of control score from the Bialer. This loading cannot be interpreted with any degree of confidence. However, such a slight positive loading does not necessarily come as a surprise since slight positive correlations between IQ and internal locus of control are generally found for heterogeneous samples (Lefcourt, 1967).

Special care should be exercised in interpreting the positive loading of religion on the g-factor given the small proportion of Protestants in the sample and parents to enroll their children in a parochial school. The principal of the school where the study was conducted indicated that most of the Protestant pupils were enrolled because they had encountered learning problems in the public schools.
Finally, the negative loadings of \( C_3 \) and \( C_4 \) on the g-factor has the effect of suggesting a slight negative relationship between the production of creative clay products and general intelligence.

Inspection of the configuration of variables loading the c-factor suggests that it defines a higher order creativity dimension analogous to general intelligence. Perhaps the term cognitive style is more appropriate (See Spotts & Mackler, 1967). In any case, the c-factor is clearly placed in the creativity domain by positive loadings from all give creativity subtests and the six creativity criteria. When considered in the framework of Campbell and Fiske (1959), the configuration of variables loading the c-factor constitutes formidable support for the convergent validity of the W-K operations.

The configuration of variables loading the c-factor suggests a tentative definition of general creativity. That is, the W-K operations can be used to define associated aspects of criterion performance. From this perspective it seems reasonable to define general creativity as a pervasive tendency to generate a plentiful supply of associations to both verbal and visual stimuli.

From a theoretical standpoint the c-factor can be readily interpreted within Wallach's (1967) framework. In this treatise Wallach described creativity in terms of the "...generation or production of ideational possibilities..." or more briefly stated, the expression of possibilities (p. 46).
### Table 8.2: Factor c - General Creativity

<table>
<thead>
<tr>
<th></th>
<th>Description</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>X1</td>
<td>Instances</td>
<td>.29</td>
</tr>
<tr>
<td>X2</td>
<td>Alternative Uses</td>
<td>.24</td>
</tr>
<tr>
<td>X3</td>
<td>Similarities (W-K)</td>
<td>.39</td>
</tr>
<tr>
<td>X4</td>
<td>Pattern Meanings</td>
<td>.41</td>
</tr>
<tr>
<td>X5</td>
<td>Line Meanings</td>
<td>.44</td>
</tr>
<tr>
<td>X14</td>
<td>Block Design</td>
<td>.23</td>
</tr>
<tr>
<td>X15</td>
<td>Object Assembly</td>
<td>.21</td>
</tr>
<tr>
<td>X25</td>
<td>Birth Order</td>
<td>-.29</td>
</tr>
<tr>
<td>C1</td>
<td>Originality (Set I-C.D.)</td>
<td>.20</td>
</tr>
<tr>
<td>C2</td>
<td>Originality (Set II-C.D.)</td>
<td>.29</td>
</tr>
<tr>
<td>C3</td>
<td>Originality (C.P.)</td>
<td>.44</td>
</tr>
<tr>
<td>C4</td>
<td>Effect. of Expression (C.P.)</td>
<td>.49</td>
</tr>
<tr>
<td>C5</td>
<td>Effect. of Expression (Set I-C.D.)</td>
<td>.34</td>
</tr>
<tr>
<td>C6</td>
<td>Effect. of Expression (Set II-C.D.)</td>
<td>.29</td>
</tr>
</tbody>
</table>
Wallach sees the generation of alternative solution strategies--especially the construction of alternative conceptual systems--as the essence of the creative act.

The birth order variable also loaded the c-factor negatively thus indicating a slight tendency for older children in a given family or children in a small family to be higher in general creativity.

Finally, two WISC subtests loaded the c-factor positively. These were the Object Assembly and Block Design subtests. The data at hand are obviously inadequate to determine just how general creativity is involved in performance on these subtests.

The two variables loading Factor I most heavily are the Line Meanings and Pattern Meanings of the W-K. Both subtests employ visual stimuli to which the examinee is instructed to provide verbal associations. Thus, Factor I may be operationally defined as the tendency to generate a plentiful supply of verbal associations to visual stimuli.

Mednick (1962) suggested the term visualizers to describe persons excelling in the generation of visual associations. With regard to the stimulus dimension Factor I seems to represent the tendency to generate a plentiful supply of verbal associations to visual stimuli. Consequently, this factor was named visual creativity or Vis. c for short.

Two complementary hypotheses may be suggested to
<p>| | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>X1:</td>
<td>Instances</td>
<td>.32</td>
</tr>
<tr>
<td>X2:</td>
<td>Alternative Uses</td>
<td>.23</td>
</tr>
<tr>
<td>X3:</td>
<td>Similarities</td>
<td>.37</td>
</tr>
<tr>
<td>X4:</td>
<td>Pattern Meanings</td>
<td>.64</td>
</tr>
<tr>
<td>X5:</td>
<td>Line Meanings</td>
<td>.68</td>
</tr>
<tr>
<td>X18:</td>
<td>Chronological Age</td>
<td>.20</td>
</tr>
<tr>
<td>X19:</td>
<td>Sex</td>
<td>.22</td>
</tr>
<tr>
<td>X25:</td>
<td>Birth Order</td>
<td>-.45</td>
</tr>
<tr>
<td>C1:</td>
<td>Originality (Set I-C.D.)</td>
<td>.45</td>
</tr>
<tr>
<td>C3:</td>
<td>Originality (C.P.)</td>
<td>.56</td>
</tr>
<tr>
<td>C4:</td>
<td>Effect of Expression (C.P.)</td>
<td>.58</td>
</tr>
</tbody>
</table>
explain the loadings of the three verbal W-K subtests on the Vis. c-factor. First, the examinee is required to transform his associations to the visual stimuli to words before the examiner can become aware of them. From this standpoint the verbal subtest may load the Vis. c-factor because they reflect fluency in transposing visual percepts into words. Second, some children may experience their associations directly in verbal form without the intermediate step of transposing for visual percept to verbal construct. According to this interpretation the verbal subtests load the Vis. c-factor because they represent the tendency to generate associations to visual stimuli in words rather than visual percepts.

The pattern of control variables loading the Vis. c-factor suggest a set of hypotheses which might be explored in further research. The negative loading by the birth order variable suggests that there is a strong tendency for older children within a family and/or children from smaller families to excel in visual creativity. The positive loading by sex of pupils suggests a tendency for boys to score higher than girls on tasks requiring visual creativity.

Finally, the strong positive loadings of C3 and C4 on the Vis. c-factor indicate that visual creativity constitutes an important component for both the Originality and Effectiveness of Expression ratings of clay products. The loading of C1 also suggests the involvement of visual creativity in
Table 8.4: Factor II - Undefined

<p>| | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>X6:</td>
<td>Information</td>
<td>-.21</td>
</tr>
<tr>
<td>C5:</td>
<td>Effect of Expression</td>
<td>.27</td>
</tr>
<tr>
<td></td>
<td>(Set I - C.D.)</td>
<td></td>
</tr>
</tbody>
</table>
that criterion.

Interpretation of Factor II does not seem justified since it is loaded slightly by only two variables. Description at an operational level seems more appropriate. In this context Factor II appears to represent a bipolar tendency to score high on C5 and low on the Information subtest of the WISC and vice versa.

The three variables loading Factor I most heavily are the Instances, Alternative Uses, and Similarities subtests of the W-K. The common denominator of performance on these tasks consists of providing verbal associations to verbal stimuli. Thus, Factor I can be operationally defined as the tendency to generate a plentiful supply of verbal associations to verbal stimuli. This factor was named verbal creativity (Ver.c) in order to reflect similarity among the defining tasks on the stimulus dimension.

Slight positive loadings by the two W-K subtests employing visual stimuli tend to suggest the involvement of visual imagery. It is not difficult to see how fluency in generating relevant visual images might enhance performance on the kinds of tasks comprising the verbal subtests of the W-K. For example, fluency in forming visual images of objects and events might reasonably be expected to facilitate thinking of uses for a knife, shoe, or newspaper.

The two control variables loading the Ver. c-factor suggest an interesting hypothesis regarding the antecedents of
Table 8.5: Factor 1 - Verbal Creativity

<table>
<thead>
<tr>
<th>X1:</th>
<th>Instances</th>
<th>.60</th>
</tr>
</thead>
<tbody>
<tr>
<td>X2:</td>
<td>Alternative Uses</td>
<td>.79</td>
</tr>
<tr>
<td>X3:</td>
<td>Similarities (W-K)</td>
<td>.58</td>
</tr>
<tr>
<td>X4:</td>
<td>Pattern Meanings</td>
<td>.33</td>
</tr>
<tr>
<td>X5:</td>
<td>Line Meanings</td>
<td>.21</td>
</tr>
<tr>
<td>X23:</td>
<td>Education of Father</td>
<td>.21</td>
</tr>
<tr>
<td>X25:</td>
<td>Birth Order</td>
<td>-.51</td>
</tr>
<tr>
<td>C3:</td>
<td>Originality (C.P.)</td>
<td>.27</td>
</tr>
<tr>
<td>C4:</td>
<td>Effect of Expression (C.P.)</td>
<td>.35</td>
</tr>
</tbody>
</table>
verbal creativity. The negative loading by the birth order variable suggests a tendency for older children within a given family and/or children from smaller families to show substantially greater verbal creativity than their classmates. The positive loading for education of father indicates that children with more highly educated fathers tend to show slightly greater verbal creativity. These findings suggest that verbal creativity is cultivated when a child has greater opportunity to engage in conversation with his parents and receives appropriate reinforcement from them.

Finally, the positive loadings by $C_3$ and $C_4$ suggest that verbal creativity constitutes a small but systematic component of criterion ratings for both dimensions of clay products.

The three WISC subtests loading Factor 2 all require the ability to determine accurately and state precisely the essential properties of a word or relationship. This ability can be most clearly observed in the scoring standards for 1 and 2 point responses on the Similarities subtest. Here a concise statement of the defining relationship is always credited with 2 points while either a less precise statement of a defining relationship or a nondefining but descriptive statement of the relationship is credited with only 1 point.

The involvement of verbal precision cannot be so clearly specified in terms of 1 or 2 point responses on the Vocabulary subtest since a 2 point response can be earned for
### Table 8.6: Factor 2 - Verbal Precision

<table>
<thead>
<tr>
<th>X9</th>
<th>Similarities (WISC)</th>
<th>.42</th>
</tr>
</thead>
<tbody>
<tr>
<td>X10</td>
<td>Vocabulary</td>
<td>.34</td>
</tr>
<tr>
<td>X12</td>
<td>Picture Completion</td>
<td>.37</td>
</tr>
<tr>
<td>X18</td>
<td>Chronological Age</td>
<td>-.39</td>
</tr>
<tr>
<td>X22</td>
<td>Religion</td>
<td>.57</td>
</tr>
<tr>
<td>X24</td>
<td>Education of Mother</td>
<td>-.31</td>
</tr>
</tbody>
</table>
reasons other than verbal precision, e.g., giving a major use, giving several descriptive features, or a correct synonym. However, 2 point responses resulting from specification of definitive features of the word can be attributed to verbal precision.

The involvement of verbal precision in performance on Picture Completion subtest can be understood in terms of the definitions offered above. Here the examinee must be able to determine and specify an essential element missing in each picture. Verbal precision is involved in the determination and description of missing elements. When the child voluntarily points to the missing element verbal precision is involved only to the extent that the examiner requires verbal responses in satisfying himself that the child has correctly identified the missing element.

For purposes of the present study, the verbal precision factor will be defined as the ability to determine accurately and state precisely the defining properties of a class, concept, or relationship. At a more abstract level it refers to the ability to employ language accurately and precisely.

The control variables loading the verbal precision factor suggest several interesting research hypotheses. First, the negative loading for C.A. is most parsimoniously explained in terms of inferior performance of the overage-in-grade pupils.
This observation suggests three alternative hypotheses. One hypothesis would be that low ability in this area may have been an important constraint to achievement in the primary grades. An alternative might be that poor school achievement resulted in a low score. Finally, some combination of influences such as cognitive style, motivation, and work-study habits may have caused both the poor school achievement and low scores on the verbal precision factor.

The verbal precision factor can be related to Cohen's (1959) oblique "Verbal Comprehension" factors. However, the interpretation offered above differ sharply from those of Cohen (1959, pp. 286-288). The correlation between Cohen's "Verbal Comprehension" (I and II) factors was 0.82 for the age group nearest to that of the present sample which caused him to acknowledge the hazards associated with any attempt at differential interpretation. Cohen did, however, suggest the possibility that "Verbal Comprehension I" consists of "...that aspect of verbally retained knowledge impressed by formal education..." (p. 287) while "Verbal Comprehension II" consists of "...the application of judgement to situations following some implicit verbal manipulation..." (p. 388).

The most straightforward interpretation of this difference in results seems to be that the hierarchial solution removed the overlap between Cohen's two oblique factors and projected it on the g-factor. The verbal precision factor of the present study would, according to this hypothesis, re-
present the nonoverlapping residuals of Cohen's two factors projected on a new reference vector.

High loadings by the Digit Span and Arithmetic subtests on Factor 3 immediately suggest that this factor corresponds to Cohen's (1959) "Freedom From Distractibility" factor. Cohen (1957, 1959) provides a penetrating analysis of the various interpretations of this factor. Among the interpretations discussed and evaluated are: memory, attention-concentration, concentration-speed, and freedom from distractibility. Cohen (1952, p. 451) defined the Freedom From Distractibility factor as follows:

(Freedom from Distractibility) A conative factor which makes it possible for problem elements to 'register' and to be maintained without loss in the course of manipulation, i.e., the ability to attend or concentrate.

The freedom from distractibility interpretation was first adopted by Cohen (1952), later rejected in favor of a memory interpretation (1957), and again adopted (1959). Cohen found it necessary to revert back to his original interpretation since this factor was loaded by the Mazes, Picture Arrangement, and Object Assembly subtests at various age levels in his factor analysis of the WISC. These three subtests clearly do not involve memory but, as Cohen (1959, p. 288) notes, are probably vulnerable to distractibility.

In the present study the "Freedom From Distractibility" factor was also loaded positively by the Comprehension, Similarities, and Vocabulary subtests of the WISC. The magnitude of their
Table 8.7: Factor 3 - Freedom From Distractibility

| X1:      | Instances            | - .22 |
| X2:      | Alternative Uses     | - .20 |
| X7:      | Comprehension        | .20   |
| X8:      | Arithmetic           | .47   |
| X9:      | Similarities (WISC)  | .23   |
| X10:     | Vocabulary           | .23   |
| X11:     | Digit Span           | .62   |
| X12:     | Picture Completion   | - .24 |
| X19:     | Chronological Age    | - .26 |
| X23:     | Education of Father  | .56   |
| X24:     | Education of Mother  | .63   |
respective loadings was slight and approximately equal which suggests a slight tendency for less distractible children to earn higher scores on verbal subtests scored for level of abstraction and/or elaboration (1 or 2) as opposed to pass-fail. This finding probably reflects a tendency for less distractible pupils to either think in more abstract terms or to consider an answer more carefully before expressing it.

Negative loadings by the Picture Completion subtest of the WISC as well as the Instances and Alternative Uses subtests of the W-K tend to suggest that a slight degree of distractibility may actually facilitate some aspects of cognition. More specifically, the "Freedom From Distractibility" factor may represent some segment of a behavioral continuum ranging from total rigidity to total flexibility. Such a hypothesis would indicate a curvilinear relationship between cognitive efficiency and the rigidity-flexibility continuum. With this orientation research might well be focused on determining the maxima with respect to the rigidity-flexibility continuum for the functions of this dimension and various cognitive tasks.

Such an interpretation can also explain the positive loadings for the three verbal WISC subtests previously discussed. That is, the positive loadings for these subtests would be explained by the fact that the maxima points for their respective functions fall further toward the rigidity end
of the continuum.

Finally, the three control variables loading this factor provide valuable data for further broadening Cohen's (1952, 1959) interpretation. The strong positive loadings by both education of father and education of mother suggest two hypotheses. First, it is possible that more highly educated parents employ socialization procedures which assist their children in learning to screen out extraneous influences and concentrate on only the relevant aspects of a stimulus configuration. According to this interpretation more educated parents tend to teach their children a task-oriented approach enabling them to resist distraction.

The second interpretation involves the possibility of hereditary factors. Since distractibility is an important component of virtually all educational disabilities, it is possible that a high proportion of the parents were themselves learning disabled which led to early termination of school attendance. This interpretation receives support from the fact that C.A. loads this factor negatively thus suggesting a substantial number of overage-in-grade pupils scored low on the "Freedom From Distractibility" factor.

Positive loadings by all four creativity criteria derived from crayon drawings suggests that Factor 4 may be defined in terms of area specific creativity. From an operational perspective, this factor represents a tendency to receive higher ratings on both the Originality and
Table 8.8: Factor 4 - Area Specific Creativity

<table>
<thead>
<tr>
<th></th>
<th>Description</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>X9</td>
<td>Similarities (WISC)</td>
<td>-.24</td>
</tr>
<tr>
<td>X13</td>
<td>Picture Arrangement</td>
<td>-.22</td>
</tr>
<tr>
<td>X18</td>
<td>Chronological Age</td>
<td>.32</td>
</tr>
<tr>
<td>C1</td>
<td>Originality (Set I - C.D.)</td>
<td>.20</td>
</tr>
<tr>
<td>C2</td>
<td>Originality (Set II - C.D.)</td>
<td>.44</td>
</tr>
<tr>
<td>C5</td>
<td>Effect of Expression</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(Set I - C.D.)</td>
<td>.31</td>
</tr>
<tr>
<td>C6</td>
<td>Effect of Expression</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(Set II - C.D.)</td>
<td>.59</td>
</tr>
</tbody>
</table>
Effectiveness of Expressions dimension of crayon drawings. It also is interesting to note that the overage-in-grade pupils tended to score somewhat higher on this factor. Also, the ratings for both dimensions of Set II loaded this factor more strongly than the ratings for Set I.

These findings suggest two considerations for future creativity research. First, creative performance may be much more specific to some areas than to others. The extent to which the three creativity dimensions (c, Vis. c, and Ver. c) are involved in criterion performance may vary widely from one area to another. For example, the results of the present study indicate that for clay products criterion performance is largely determined by the three factors discussed above while in the case of crayon drawings task specific considerations assume more importance.

Second, the operations used in defining a particular criterion dimension may reasonably be expected to affect the factorial composition of performance measurements for a given area unless that area is unitary and highly specific. For example, one would not necessarily expect the factorial composition of Beittel's (1964) process criterion to be equivalent to that of the Martinson-Seagoe (1967) dimensions for clay products.

Examination of the configuration of variables loading Factor 5 suggest that this factor can best be understood primarily in terms of the high loading by the Coding subtest.
### Table 8.9: Factor 5 - Coding

<table>
<thead>
<tr>
<th>Variable</th>
<th>Description</th>
<th>Correlation</th>
</tr>
</thead>
<tbody>
<tr>
<td>X3</td>
<td>Similarities (W-K)</td>
<td>0.27</td>
</tr>
<tr>
<td>X11</td>
<td>Digit Span</td>
<td>0.21</td>
</tr>
<tr>
<td>X16</td>
<td>Coding</td>
<td>0.60</td>
</tr>
<tr>
<td>X19</td>
<td>Sex</td>
<td>-0.68</td>
</tr>
</tbody>
</table>
The strong loading by the Coding subtest suggests the present factor may bear some similarity to Cohen's "Quasi-Specific" factor (Cohen, 1959, pp. 287-288). However, the present factor differs from Cohen's factor in three ways. First, the Coding subtest loaded the present factor much more strongly than the "Quasi-Specific" factor was loaded by the Coding in Cohen's study. Second, the Picture Arrangement subtest failed to load the present factor as it did Cohen's factor. Third, the slight positive loading by the Digit Span subtest was not obtained by Cohen. These differences suggest that a direct correspondence cannot necessarily be established between the present factor and that of Cohen.

The strong negative loading by sex of pupil indicates that in the present sample girls showed a strong tendency to score higher than boys on the Coding subtest. This finding stands in sharp contrast to the standardization data reported by Wechsler (1949). The discrepancy between the results of the present study and Wechsler's data should be systematically explored. That is, the performance of boys and girls on the Coding subtest should be compared to determine if the results of the present study reflect some specific characteristics of the sample. The alternative explanation is, of course, that Wechsler's standardization data is inaccurate with regard to sex differences.

The slight positive loading by the Digit Span sub-
test suggests the involvement of short-term memory processes. Such a hypothesis seems reasonable in that the child who readily commits the number-symbol combination to memory should be able to perform better than the child who must frequently refer back to the key.

The slight positive loading by the Similarities sub-test of the W-K suggests the involvement of motivational elements. Superior performance on both the Similarities subtest of the W-K and the Coding subtest requires sustained effort beyond that necessary for meeting the minimum task requirements. For example, the child who gives two or three associations to the Similarities subtest or completes the Coding subtest has satisfied the task requirements. The child who strives for more effective performance apparently does so because of strong personal involvement. As Glasser and Zimmerman (1967, pp. 93-95) have indicated, such motivational influences may range from outright compulsiveness to task flexibility.

Thus, the results of the present analysis suggest that performance on the Coding subtest is most directly determined by visual-motor coordination while short-term memory and emotional influences function as secondary determiners. The possibility of a strong sex difference on this subtest is also suggested since girls earned considerably higher scores than boys.

Examination of the configuration of variables loading Factor 6 suggests that this factor cannot be clearly associated
Table 8.10: Factor 6 - Intact Home

<p>| | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>X4:</td>
<td>Pattern Meanings</td>
<td>-.22</td>
</tr>
<tr>
<td>X9:</td>
<td>Similarities (WISC)</td>
<td>.24</td>
</tr>
<tr>
<td>X11:</td>
<td>Digit Span</td>
<td>.22</td>
</tr>
<tr>
<td>X17:</td>
<td>Locus of Control (internal)</td>
<td>.23</td>
</tr>
<tr>
<td>X20:</td>
<td>Intactness of Home</td>
<td>.36</td>
</tr>
<tr>
<td>Cl:</td>
<td>Originality (Set I - C.D.)</td>
<td>.26</td>
</tr>
</tbody>
</table>
with either the intelligence or creativity domains. This factor was loaded moderately by the intactness of home factor from which it was named. Since the intact home factor is weakly defined, caution must necessarily be observed in interpretation. At an operational level, however, the findings suggest that children coming from homes with both natural parents present may be described as follows: 1) slightly higher in detecting and describing semantic relationships, 2) slightly higher in remembering series of random digits, 3) slightly more willing to attribute events to circumstances within their own control, 4) slightly lower in number of verbal association to a specific class of visual stimuli, and 5) slightly higher in Originality ratings for one set of crayon drawings.

The pattern of variables loading the Intact Home factor does suggest one tentative hypothesis. It may well be that the security generated by a stable home life exercises a pervasive influence across many facets of both cognitive and personality development. One might go even further and surmise that the basic dimension involved here is the influence of a stable family situation rather than an intact home per se. Such a hypothesis is congruent with the moderate loading by the intactness of home variable which clearly indicates that not all intact homes produce children scoring high on this factor.

The overall pattern of variables loading this factor is indicative of bipolarity. In operational terms, lack of
Table 8.11: Factor 7 - Education of Father

<table>
<thead>
<tr>
<th>X1:</th>
<th>Instances</th>
<th>.24</th>
</tr>
</thead>
<tbody>
<tr>
<td>X2:</td>
<td>Alternative Uses</td>
<td>.25</td>
</tr>
<tr>
<td>X8:</td>
<td>Arithmetic</td>
<td>.31</td>
</tr>
<tr>
<td>X9:</td>
<td>Similarities (WISC)</td>
<td>-.29</td>
</tr>
<tr>
<td>X18:</td>
<td>Chronological Age</td>
<td>.20</td>
</tr>
<tr>
<td>X23:</td>
<td>Education of Father</td>
<td>-.51</td>
</tr>
<tr>
<td>C2:</td>
<td>Originality (Set II - C.D.)</td>
<td>.25</td>
</tr>
<tr>
<td>C4:</td>
<td>Effect of Expression (C.P.)</td>
<td>.20</td>
</tr>
</tbody>
</table>
education on the part of the father tends to be associated with the following: 1) slightly lower scores on the Similarities subtest, 2) slightly higher scores on the Arithmetic subtest, 3) slightly higher scores on two verbal subtests of the W-K, 4) slightly higher ratings on two creativity criteria, and 5) a slight tendency to be over-age in grade. Children with more highly educated fathers would, of course, show behaviors opposite those described above.

As with the previous factor, one is tempted to surmise that the Factor 7 can be explained in terms of the pervasive influence of a control variable. That is, one is tempted to advance the hypothesis that more highly educated fathers behave in such a manner as to foster a particular pattern of cognitive development.

Extreme caution is, however, necessary in going beyond this point and implicitly concluding that educational attainment causes the father to behave in a given manner. Instead, it is quite possible that certain personality characteristics such as determination or dominance functioned to facilitate the acquisition of advanced education on the part of the father and foster a particular pattern of cognitive development on the part of his children.

The configuration of variables loading Factor 8 shows considerable similarity to the "Perceptual Organization" factor described by Cohen (1957, 1959). Cohen (1959, p. 287) observed that this factor was defined by loadings from the Object Assembly,
| X8:    | Arithmetic   | .27 |
| X12:   | Picture Completion | .22 |
| X14:   | Block Design   | .66 |
| X15:   | Object Assembly | .28 |
| X17:   | Locus of Control (internal) | .25 |
| C5:    | Effect of Expression (Set I - C.D.) | .36 |
Block Design, and Mazes subtests for his age group closest to that of the present sample. The Mazes subtest was not administered in the present study. However, the similarity between the present study and those of Cohen is still obvious since this factor has been consistently defined by positive loadings by the Block Design and Object Assembly subtests at all age levels (Cohen, 1957, 1959). Cohen described the tests loading this factor as "all non-verbal and require the interpretation and/or organization of visually perceived materials against a time limit" (1959, p. 287).

However, the results of the present study suggest a broadening of Cohen's interpretation. In specific, the additional variables included in the present analysis suggest that this factor may consist of a basic dimension of human cognition which subsumes effective organization of a complex array of visual stimuli.

Such a hypothesis is supported by the findings of Mischel and Metzner (1962) which showed a substantial correlation between the Block Design subtest and the Embedded Figures Test (EFT). Witkin et al. (1962) described individuals scoring high on the EFT as field-independent. In the course of a series of studies Witkin et al. concluded that field independent persons not only have a superior ability to detach themselves from the influence of their immediate surroundings but also have a more differentiated personality structure.
Overview of the Factor Structure:

Separation between the creativity and intelligence domains is clearly evident from inspection of the hierarchical solution. Factors in the intelligence domain were defined by positive loadings from WISC subtests while those in the creativity domain were defined by positive loadings from W-K subtests and creativity criteria. The fact that these respective dimensions remained totally refractory to merger at all levels of the hierarchy constitutes strong evidence for the discriminant validity and dimensionality of the W-K (1965) creativity operations.

The hierarchical solution not only shows a clear separation between the creativity and intelligence domains but also delineates the structural relations among the various dimensions comprising these two domains. In the case of the intelligence domain the structural relations are quite clear as a strong g-factor and four primary factors were obtained. The creativity domain consists of a general creativity (c) factor but bifurcates at the subgeneral level to form a visual creativity (Vis. c) factor and a verbal creativity (Ver. c) factor. This bifurcation appears to be based on the stimulus dimension, i.e., whether the examinee is required to employ visual or verbal information processing. A task specific creativity factor defined by criterion ratings for crayon drawings emerged at the primary level.
Three more factors were obtained but they could not be assigned to either the creativity or intelligence domains. One of these was a weak bipolar factor bordering between the creativity and intelligence domains. The remaining two factors were defined by background variables.

Examination of the relationship among the basic dimensions comprising the creativity domain provides support for Fee's (1968) hypothesis. In specific, the results of the hierarchial analysis strongly suggest that creativity as defined by the W-K operations is multidimensional in nature. At the highest level one finds the general creativity (c) factor which appears to prevade creative endeavor in a manner analogous to that of g in the intelligence domain. The c-factor was defined by positive loadings from all creativity operations, i.e., six creativity criteria and five W-K subtest.

The c-factor can be interpreted from both a theoretical and operational perspective. In operational terms it represents a tendency to generate a plentiful supply of verbal associations to both verbal and visual stimuli. According to this interpretation the operations employed in the W-K subtests are used to define common elements in the criterion variables. At an even more basic level the c-factor may be operationally defined as the positive intercorrelation among all creativity assessors employed in the study.

Comparison of the g and c factors suggests that general creativity is relatively weak in relation to general
intelligence. However, this conclusion is not necessarily justified as the intelligence assessors may simply reflect more precise measurement. This position is not untenable since the creativity operations are relatively recent while the intelligence operations evolved from decades of research and development. Furthermore, the problems encountered with the creativity criteria obtained from crayon drawings necessarily limited the strength of the c-factor. Thus, it seems wise to suspend judgement rather than attempt to determine the relative strength of g and c on the basis of data obtained from the present study.

The c-factor can be interpreted within the theoretical framework of Wallach (1967) and Wallach and Wing (1969). Within the framework of Wallach and his associates the c-factor might be interpreted as the tendency to express a greater number of possibilities across a wide range of situations involving both verbal and nonverbal response media. The exposition of Wallach and Wing (1969) would suggest that the expression of possibilities is associated with "...greater cognitive energy, or a higher typical level of ideational activity, or ...greater cognitive vitality..." (p. 110).

Slight loadings by the Block Design and Object Assembly subtests of the WISC are not sufficient to disrupt the essential characteristics of the c-factor. Rather, these loadings suggest that fluency in generating possible solution strategies facilitates performance on the Block Design and Object Assembly subtests.
Involvement by the birth order variable suggests that being an older child within a given family and/or coming from a smaller family is a relatively important antecedent of ideational productivity. On the basis of this finding one is tempted to surmise that extensive parental attention during early childhood facilitates the development of ideational productivity in later childhood and perhaps beyond.

The visual creativity (Vis. c) factor was defined by strong loadings from the W-K subtests where visual stimuli are employed. Both creativity criteria obtained for clay products also loaded this factor strongly. The Originality criterion for the first set of crayon drawings also loaded moderately. However, the three W-K subtest consisting of verbal stimuli also showed slight to moderate loadings.

Two hypotheses can be advanced to explain the loadings by the three verbal subtests on the Vis. c-factor. First, the examinee is required to verbalize his associations to visual stimuli even though these associations occur in the form of visual percepts. Merely developing a visual image is not enough. The examinee must express his percept in verbal form before the examiner can become aware of it. Second, the examinee may experience his associations in verbal form without the intermediate step of transforming a visual percept to verbal form. One can easily surmise that the manner of generating percepts varies considerably among individual in such a way that both hypotheses are of value in explaining the involvement of the verbal subtests.
From an operational perspective the Vis. c-factor can be interpreted as either a tendency to generate a plentiful supply of verbal associations to visual stimuli or positive intercorrelation among nonverbal creativity tasks.

The verbal creativity (Ver. c) factor was defined by strong positive loadings from the three W-K subtests consisting of verbal stimuli. The two visual subtests of the W-K also loaded this factor thus suggesting the involvement of visualization. Such involvement is not difficult to understand since fluency in developing visual images might reasonably be expected to enhance a child's ability to name instances of a specific class, determine similarities, or think of uses for a familiar object.

From an operational perspective the Ver. c-factor can be defined as either a tendency to generate a plentiful supply of verbal associations to verbal stimuli or positive intercorrelation among verbal creativity tasks.

Bifurcation of the creativity domain into the Vis. c-factor and Ver. c-factor suggests that creativity as operationally defined by the W-K test may constitute a more complex phenomena than W-K (1965) initially suspected. More specifically, the hypotheses of Mednick (1962) and Fee (1968) were confirmed thus indicating that creativity is a multidimensional phenomena. This finding has the effect of suggesting elaboration of the W-K (1965) creativity theory beyond the efforts of Wallach (1967) and Wallach and Wing (1969). One cannot reasonably state that W-K's formulations are incorrect
since their operations showed substantial positive inter-correlations and defined an aspect of cognition independent from general intelligence. Rather, the findings of the present study and those of Fee (1968) indicate the necessity for giving consideration to the class (verbal-visual) of stimuli presented to the subject. The W-K creativity theory could be readily expanded to accommodate these findings.

Both the verbal and visual creativity factors were loaded negatively by the birth order variable. From an operational perspective these loadings indicate a substantial tendency for older children in a given family and/or children from small families to score higher on both verbal and visual creativity assessors. Again, the degree of parental attention during early childhood is suggested as an important antecedent of creative performance. It may well be that the strong negative relationship between creative performance and birth order reflects parental reinforcement patterns rather than increased parent-child interaction in a smaller family constellation may serve to increase the likelihood for ideational productivity to receive reinforcement.

In any case, birth order may have attained increased importance in the present study because of family size. Children in the present sample came from families much larger than those of suburban children who are frequently employed in creativity research.

An area specific creativity factor was defined by
positive loadings from all for creativity criteria obtained from crayon drawings. Emergence of this factor suggests, but does not necessarily establish, the importance of task-specific creativity.

The factorial composition of the W-K creativity operations seems to be relatively well established by the results of the present study and that of Fee (1968). However, these findings by no means rule out the possibility of task and/or area specific creativity in the hierarchical orientation suggested above. Quite the contrary, the fact that a creativity factor specific to crayon drawings emerged in the present study suggests the possibility of task-specific creativity.

Thus, one might approach creativity research involving a specific task or set of similar tasks with the idea of determining the proportion of criterion variance accounted for by general creativity, verbal creativity, and visual creativity. At the same time he would be aware of the possibility that a substantial proportion of the criterion variance might well be area specific.

The hierarchial solution employed in the present study shows substantial agreement with Cohen's (1959) oblique solution. Correspondence between Cohen's "Perceptual Organization", "Freedom From Distractibility", and "Quasi-Specific" factors and three primary factors from the present study is clearly evident. The verbal precision factor of the present
study probably represents the residuals of Cohen's "Verbal Comprehension I and II" factors after overlap was projected to the g-factor. This difference can be understood in methodological terms since the hierarchial solution provides for orthonogality among factors at all levels.

The results of the present study are also in agreement with those of Cohen (1959) in indicating that the Vocabulary, Information, Comprehension, and Arithmetic subtests of the WISC constitute the best estimates of g. Furthermore, both studies agree with regard to the relative magnitude of loadings by these subtest on the g-factor.

From an overall perspective, the agreement between Cohen's (1959) results and those of the present study is rather surprising when one considers the differences in method of analysis, sample, and range of variables included.

The most important discrepancy between the results of Cohen (1959) and those of the present study is to be found in the proportion of subtest variance attributed to the g-factor. However, this discrepancy can also be understood in methodological terms. Cohen obtained his estimates of subtest variance attributable to g by squaring the correlation between the various subtests and scores for the g-factor. His g-factor scores were obtained from a linear equation which provides an optimum estimate of g. However, this procedure minimizes but cannot completely eliminate variance attributable to the pri-
mary factors from the g-score. The correlation between the g-score and a given subtest is increased to the extent that both contain variance attributable to one or more primary factors.

The hierarchial solution provides a more direct means of obtaining variance estimates for g in that one needs only to square the respective subtest loadings on the g-factor. In the opinion of the investigator this procedure provides a better estimate of subtest variance attributable to g than the method employed by Cohen.

Question 4: Among the variables examined; five creativity subtests, eleven WISC subtests, and nine control variables; what is the most effective combination for predicting concurrent creative performance in the areas of visual art selected for study?

The data reported in this section were analyzed by use of the WETSL Program (Wherry, 1969) designed for use on the IBM 360. This program represents the most recent computerized version of the Wherry-Doolittle test selection method, a well established stepwise regression procedure (see Garrett, 1958). With this method, predictors are selected sequentially until maximum prediction of the population value is attained. At each step of this procedure one is provided with a multiple correlation coefficient ($R$), a shrunken multiple correlation coefficient ($R^*$), and F-ratio for the increment in the squared multiple correlation coefficient ($R^2$). In addition beta-weights (standard score), bse-weights (raw score), t-values for these weights
and a constant for the raw score equation are provided after each increment.

The \( \bar{R} \) mentioned above is provided by incorporation of the Wherry shrinkage formula (Wherry, 1931) in the WETSL program. Application of the shrinkage formula to the statistic \( R \) provides an estimate of correlation in the population. That is, \( \bar{R} \) provides an indication of the value of \( R \) which one might expect to obtain in the course of cross-validation with a similar sample.

**Clay Products**

A few comments regarding the method of presentation followed in this subsection are necessary. The writer elected to present the results of the full regression analysis rather than stop at a point indicated by conservative decision rules. This decision was influenced by the work of Taylor and Ellison (1964) who found that creativity criteria were best predicted by a large number of variables each of which made a small contribution to the multiple correlation coefficient. These researchers found that their regression equation held up exceptionally well in cross-validation.

One may well encounter the same situation when dealing with reliable criteria such as those obtained from clay products. Consequently, the constant, score weights (\( \text{bee and beta} \)), and \( t \)-values for respective weights are those for the full stepwise regression procedure. Values for the statistics noted above are tabled throughout the discussion in the text.
However, values of the above statistics determined by more conservative decision rules are presented in the Appendixes. Conservative decision rules would indicate that the regression coefficients and the constant associated therewith, should be determined from the last point in the stepwise procedure where the increment in $R^2$ attained significance at the .05 level. References to the appropriate Appendix for the weights determined in this way are contained in text.

As previously indicated, two criterion ratings were secured for clay products. That is, the judges rated the set of clay products on both the Originality and Effectiveness of Expression dimensions. The analysis of ratings on the Originality dimension ($C_3$) is treated first in this subsection. The results from analysis of the Effectiveness of Expression dimension ($C_4$) are discussed next. Finally, the regression equations for the two criteria are examined in the context of the factorial composition of the predictor and criterion variables.

Originality Ratings ($C_3$): Tables 9 and 10 provide a summary of the multiple regression procedure for $C_3$. Table 9 contains data essential to an understanding of the stepwise regression procedure. Table 10 contains the final regression for $C_3$. Appendix H contains the regression equation for $C_3$ when the stepwise procedure was terminated with the last increment significant at the 0.05 level.
Table 9: Stepwise Regression Procedure for C3

<table>
<thead>
<tr>
<th>Variable</th>
<th>5</th>
<th>.641</th>
<th>.634</th>
<th>49.391</th>
<th>1.71</th>
<th>.001</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>2</td>
<td>.669</td>
<td>.657</td>
<td>4.750</td>
<td>2.70</td>
<td>.05</td>
</tr>
<tr>
<td></td>
<td>17</td>
<td>.682</td>
<td>.665</td>
<td>2.307</td>
<td>3.69</td>
<td>.10</td>
</tr>
<tr>
<td></td>
<td>11</td>
<td>.691</td>
<td>.669</td>
<td>1.618</td>
<td>4.68</td>
<td>.25</td>
</tr>
<tr>
<td></td>
<td>6</td>
<td>.704</td>
<td>.677</td>
<td>2.323</td>
<td>5.67</td>
<td>.10</td>
</tr>
<tr>
<td></td>
<td>18</td>
<td>.718</td>
<td>.686</td>
<td>2.673</td>
<td>6.66</td>
<td>.05</td>
</tr>
<tr>
<td></td>
<td>1</td>
<td>.726</td>
<td>.690</td>
<td>1.591</td>
<td>7.65</td>
<td>.25</td>
</tr>
<tr>
<td></td>
<td>19</td>
<td>.733</td>
<td>.692</td>
<td>1.379</td>
<td>8.64</td>
<td>.25</td>
</tr>
<tr>
<td></td>
<td>10</td>
<td>.740</td>
<td>.695</td>
<td>1.580</td>
<td>9.63</td>
<td>.25</td>
</tr>
<tr>
<td></td>
<td>16</td>
<td>.746</td>
<td>.697</td>
<td>1.213</td>
<td>10.62</td>
<td>.50</td>
</tr>
</tbody>
</table>
Table 10: Regression Equations for C4

<table>
<thead>
<tr>
<th>Variable</th>
<th>Beta-Weight</th>
<th>Bee-Weight</th>
<th>t for Weight</th>
</tr>
</thead>
<tbody>
<tr>
<td>5</td>
<td>0.533</td>
<td>0.224</td>
<td>5.094</td>
</tr>
<tr>
<td>2</td>
<td>0.047</td>
<td>0.028</td>
<td>0.335</td>
</tr>
<tr>
<td>17</td>
<td>-0.134</td>
<td>-0.490</td>
<td>-1.462</td>
</tr>
<tr>
<td>11</td>
<td>0.144</td>
<td>0.506</td>
<td>1.523</td>
</tr>
<tr>
<td>6</td>
<td>-0.167</td>
<td>-0.678</td>
<td>-1.519</td>
</tr>
<tr>
<td>18</td>
<td>-0.210</td>
<td>-0.203</td>
<td>-2.045</td>
</tr>
<tr>
<td>1</td>
<td>0.175</td>
<td>0.052</td>
<td>1.214</td>
</tr>
<tr>
<td>19</td>
<td>0.155</td>
<td>2.977</td>
<td>1.676</td>
</tr>
<tr>
<td>10</td>
<td>-0.137</td>
<td>-0.534</td>
<td>-1.221</td>
</tr>
<tr>
<td>16</td>
<td>0.110</td>
<td>0.479</td>
<td>1.101</td>
</tr>
</tbody>
</table>

A-Weight: 38.633
Analysis of the regression of the criterion on the predictor clearly indicated that the Line Meanings subtest (X5) of the W-K is the most effective single predictor of ratings for the Originality dimension of clay products. The correlation between X5 and C3 was 0.641 which attained significance at beyond the 0.001 level. This predictor alone explained approximately 41 per cent of the criterion variance. Thus, the Line Meanings subtest of the W-K is sufficient to account for a high proportion of variance in the Originality dimension of clay products.

The second predictor selected was also one of the W-K subtests, Selection of the Alternative Uses subtest (X2) resulted in an increment in $R^2$ significant at the 0.05 level. The Alternative Uses subtest, in contrast with the Line Meanings subtest, requires verbal association to verbal stimuli.

Internal locus of control (X17) as measured by the Bialer was the third variable selected. Addition of X17 resulted in an increment in $R^2$ significant at only the 0.10 level. The regression weight for X17 was, however, negative thus indicating the subtraction of element from C3 not positively related to the predictors already selected.

The Digit Span subtest (X11) of the WISC was next selected because of a slight positive relationship with the criterion. However, addition of X11 resulted in only a slight increment which attained significance at only the 0.25 level.

The next variable selected was the Information sub-
test (X6) of the WISC. Addition of X6 resulted in an increment significant at beyond the 0.10 level. The Information subtest functioned to subtract the effects of the general intelligence factor (g) from the criterion. Alternatively, selection of X6 is indicative of the tendency for pupils with high criterion ratings to earn slightly lower scores on WISC subtests highly saturated with g. Selection of X6 removed the effects of this trend from the criterion.

With the effects of g removed, chronological age (X18) was selected. The increment resulting from addition of X18 to the equation attained significance at beyond the 0.05 level. Selection of X18 functioned to subtract the effects of age from the criterion. That is, the tendency for over-age pupils in a given grade to receive slightly lower criterion ratings was removed. The effects of age were, of course, removed from the predictors as well as from the criterion. In fact, removal of the age effects from the predictors selected earlier in the analysis resulted in a considerable reshuffling of beta-weights.

The most obvious effect of removing the age effect was the selection of the Instances subtest (X1) of the W-K and an accompanying drop of the beta-weight for X2. In fact, the decrease in the beta weight for X2 was so great that it could probably be dropped from the equation without noticeable loss of precision. The weight for X6 was sharply increased thus indicating increased importance of the g-factor with the
age effect fully removed.

The following three variables were also selected but their importance in the regression equation was relatively limited: sex of pupil \((X_{19})\), Vocabulary \((X_{10})\), and Coding \((X_{16})\).

The linear regression equation provided by the stepwise analysis resulted in an \(R\) of 0.746. Thus, the ten predictors included in the equation summarized in Table 10 account for about 56 percent of the criterion variance.

The degree of relationship between the ten predictors and criterion was highly significant. An \(F\)-ratio of 7.78 with df \((10,62)\) was obtained. This value attained significance at beyond the 0.001 level.

Application of the Wherry shrinkage formula resulted in an \(R\) of 0.697 thus indicating that one might reasonably expect the regression equation reported in Table 10 to account for about 49 per cent of the criterion variance in a cross-validation study with a similar sample. The extent to which this equation can be generalized to pupils with diverse characteristics is, of course, an empirical question to be answered by further research.

The regression equation described above can best be understood in the context of the factorial composition of the criterion and the predictor variables. This is especially true with regard to the W-K subtests included in the equation. The prominence of the Line Meanings subtest \((X_{5})\) in the equation
can be understood in terms of c-factor and Vis. c-factor. In
specific, these two factors were important components of effec-
tive performance on both the Line Meanings subtest and the cri-
terion.

It is also likely that the Ver. c-factor was involved
to some degree in the selection of the Instance subtest \( (X_1) \).
However, the involvement of the Ver. c-factor is necessarily
limited since the criterion loaded this factor to only a slight
degree.

The role of some of the WISC subtests in the equation
can also be specified in terms of their factorial composition.
In specific, the negative loading of the criterion on the f-fac-
tor suggests a slight negative relationship between general in-
telligence and Originality ratings for clay products. Given
such a relationship the role of the Information \( (X_6) \) and Voca-
bulary \( (X_{10}) \) subtests in the equation can be understood since
these two subtests load the g-factor stronger than any of the
other predictors.

On the other hand, the involvement of the Digit Span
\( (X_{11}) \) and Coding \( (X_{16}) \) subtests cannot be determined with any
degree of confidence from inspection of the hierarchial factor
matrix. In the absence of more definitive data these subtests
are best regarded as weak positive correlates of Originality
ratings for clay products. Thus, one is left with the sugges-
tion that persons scoring high on the criterion also tend to
score slightly higher on the Coding and Digit Span subtests of
the WISC.

A similar conclusion is necessary for internal locus of control ($X_{17}$). The role of this predictor cannot be determined since it does not load strongly on any of the same factors as the criterion. Here again, one is justified only in surmising that internal locus of control bears a slight negative relationship to criterion performance.

The situation is somewhat more promising with regard to interpretation of sex of pupil ($X_{19}$). At an empirical level the positive weight for this predictor indicates a tendency for boys to receive slightly higher criterion ratings than girls. This finding may well be accounted for by the tendency for boys to score slightly higher than girls on the Vis. c-factor which is a major component of criterion performance.

Finally, the prominence of chronological age ($X_{18}$) in the equation is noteworthy in that it clearly demonstrates the importance of control variables in creativity research. The potential consequences of failure to control for age in the present study are quite apparent. Such an omission would have led to less efficient prediction and, at the same time, have obfuscated the relationship between the g-factor and criterion performance.

The regression equation can also be considered from an alternative vantage point. When considered in relation to the results of the factor analysis, the regression equation can be used to help define the elements involved in criterion
performance. Since the operations employed in the predictors can be clearly established they can in turn be used to define criterion performance if two conditions are met. First, a given predictor must account for a substantial proportion of the criterion variance. Second, both the predictor and criterion must show substantial loadings on a very few (seldom more than one or two) factors which can be defined in terms of the known predictor variable.

Both these conditions were clearly met in the case of the Line Meanings subtest of the W-K. Consequently, this subtest can be used to describe the elements involved in Originality ratings for clay products. The c-factor and Vis. c-factor were clearly involved. Thus, it can be reasonably argued that the tendency to generate a large number of verbal associations to visual stimuli and the tendency to generate a large number of verbal associations to both verbal and visual stimuli constitute important components of criterion ratings.

In summary, the data discussed above provide a tentative answer to the research question considered in this subsection. The extent to which the present regression can be generalized to other samples is an empirical matter which can be resolved only by further research. Insofar as the present sample is concerned the general hypothesis was generally confirmed. That is, the W-K creativity operations constituted the most effective class of predictor variables for C3 but maximum prediction was achieved by including both intelligence and
control variables. From a conceptual standpoint the role of the creativity subtests was much clearer than that of either the intelligence or control variables. That is, the importance of the W-K operations can be understood in terms of the involvement of the c-factor, Vis. c-factor, and perhaps slight involvement of the Ver. c-factor in these subtests and the Originality criterion.

Interpretation of the intelligence and control variables in the equation poses a much more difficult problem. Several considerations figure in this difficulty. First, the majority of the variables account for only slight increments. When this occurs it becomes increasingly difficult to discriminate between weak relationships and chance factors. Second, the suppressor relationships frequently observed among the intelligence and control variables rendered interpretation extremely hazardous. Third, the involvement of these variables in the equation could not be effectively determined from the results of the factor analysis. Given these constraints it seemed prudent to offer only the most guarded interpretations for most of these variables in the equation. In some cases variables were simply allowed to stand as slight correlates of criterion performance with no interpretation.

**Effectiveness of Expression Ratings (C₄):** Table 11 and 12 provide a summary of the multiple regression procedure for C₄. Table 11 summarizes the stepwise regression procedure while Table 12 provides both the standard and raw score forms
Table 11: Stepwise Regression Procedure for C4

<table>
<thead>
<tr>
<th>Variable</th>
<th>R</th>
<th>R̄</th>
<th>F</th>
<th>df</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>5</td>
<td>.669</td>
<td>.663</td>
<td>57.573</td>
<td>1.71</td>
<td>.001</td>
</tr>
<tr>
<td>2</td>
<td>.722</td>
<td>.713</td>
<td>10.822</td>
<td>2.70</td>
<td>.001</td>
</tr>
<tr>
<td>7</td>
<td>.745</td>
<td>.732</td>
<td>5.095</td>
<td>3.69</td>
<td>.01</td>
</tr>
<tr>
<td>19</td>
<td>.752</td>
<td>.735</td>
<td>1.814</td>
<td>4.68</td>
<td>.25</td>
</tr>
<tr>
<td>17</td>
<td>.759</td>
<td>.738</td>
<td>1.621</td>
<td>5.67</td>
<td>.25</td>
</tr>
<tr>
<td>4</td>
<td>.765</td>
<td>.740</td>
<td>1.431</td>
<td>6.66</td>
<td>.25</td>
</tr>
<tr>
<td>20</td>
<td>.775</td>
<td>.745</td>
<td>2.579</td>
<td>7.65</td>
<td>.05</td>
</tr>
<tr>
<td>8</td>
<td>.781</td>
<td>.749</td>
<td>1.446</td>
<td>8.64</td>
<td>.25</td>
</tr>
<tr>
<td>6</td>
<td>.791</td>
<td>.757</td>
<td>2.756</td>
<td>9.63</td>
<td>.01</td>
</tr>
<tr>
<td>25</td>
<td>.799</td>
<td>.761</td>
<td>1.951</td>
<td>10.62</td>
<td>.10</td>
</tr>
<tr>
<td>9</td>
<td>.808</td>
<td>.768</td>
<td>2.587</td>
<td>11.61</td>
<td>.01</td>
</tr>
<tr>
<td>11</td>
<td>.812</td>
<td>.769</td>
<td>1.111</td>
<td>12.60</td>
<td>.50</td>
</tr>
</tbody>
</table>
Table 12: Regression Equation for C4

<table>
<thead>
<tr>
<th>Variable</th>
<th>Beta-Weight</th>
<th>Bee-Weight</th>
<th>t for Weight</th>
</tr>
</thead>
<tbody>
<tr>
<td>5</td>
<td>0.769</td>
<td>0.318</td>
<td>5.133</td>
</tr>
<tr>
<td>2</td>
<td>0.273</td>
<td>0.160</td>
<td>2.621</td>
</tr>
<tr>
<td>7</td>
<td>-0.150</td>
<td>-0.605</td>
<td>-1.662</td>
</tr>
<tr>
<td>19</td>
<td>0.111</td>
<td>2.096</td>
<td>1.345</td>
</tr>
<tr>
<td>17</td>
<td>-0.148</td>
<td>-0.534</td>
<td>-1.783</td>
</tr>
<tr>
<td>4</td>
<td>-0.422</td>
<td>-0.234</td>
<td>-2.369</td>
</tr>
<tr>
<td>20</td>
<td>-0.170</td>
<td>-3.181</td>
<td>-2.073</td>
</tr>
<tr>
<td>8</td>
<td>0.150</td>
<td>0.487</td>
<td>1.470</td>
</tr>
<tr>
<td>6</td>
<td>-0.131</td>
<td>-0.523</td>
<td>-1.430</td>
</tr>
<tr>
<td>25</td>
<td>-0.227</td>
<td>-1.308</td>
<td>-1.999</td>
</tr>
<tr>
<td>9</td>
<td>-0.176</td>
<td>-0.696</td>
<td>-1.864</td>
</tr>
<tr>
<td>11</td>
<td>0.103</td>
<td>0.358</td>
<td>1.054</td>
</tr>
</tbody>
</table>

A-Weight: 32,911
of the final regression equation. Appendix I contains the regression equations for $C_4$ when the stepwise procedure was terminated in accordance with the more conservative decision rule.

Analysis of the regression of the criterion on the predictor variables revealed the importance of the Line Meanings subtest ($X_5$). The correlation between this predictor and the criterion was 0.669 which attained significance at beyond the 0.001 level. Thus, it can be readily observed that the Line Meanings subtest alone accounted for 45 per cent of the criterion variance.

The second variable selected was also one of the W-K subtests. Addition of the Alternative Uses subtest ($X_2$) to the regression equation resulted in an increment significant at beyond the 0.001 level. An $R$ of 0.722 was obtained at this point which indicates that these two W-K subtests account for approximately 52 per cent of the criterion variance.

The Comprehension subtest ($X_7$) was next selected with a negative regression weight. The increment in $R^2$ resulting from addition of $X_7$ attained significance at beyond the 0.01 level thus indicating a substantial contribution to the equation. This subtest obviously functioned to subtract some as yet undefined aspect of criterion performance. The nature of the elements removed cannot, however, be defined from knowing only that they are negatively related to the Comprehension subtest.

The next three variables added to the equation were
sex of pupil \((X_{19})\), internal locus of control \((X_{17})\), and the Pattern Meanings subtest \((X_4)\) of the W-K. Each of these variables resulted in a slight increment in \(R^2\) which attained significance at only the 0.25 level. However, each variable functioned in a different manner in the equation.

The control variable sex of pupil \((X_{19})\) entered into the equation with a positive weight thus indicating a slight tendency for boys to receive higher criterion ratings than girls. Examination of the hierarchial factor matrix suggests involvement of the Vis. c-factor. That is, the tendency for boys to earn superior scores on the Vis. c-factor may account for the role of \(X_{19}\) in the equation.

The involvement of internal locus of control \((X_{17})\) in the equation is clearly one of subtracting criterion variance not measured by the remaining predictors. The nature of the influence removed from the criterion cannot be described with any degree of confidence with only a knowledge of the Bialer from which \(X_{17}\) was obtained. This is especially true when one considers the factorial complexity of \(X_{17}\). Given these considerations, internal locus of control is best regarded as a negative correlate of criterion performance.

The Pattern Meanings subtest \((X_4)\) of the W-K functioned in a highly complex manner with regard to the remainder of the regression equation. This variable entered the equation with a negative regression weight at the sixth step of the analysis. The absolute magnitude of the beta-weight for \(X_4\) increased dur-
ing the final six steps of the analysis until it attained considerable prominence in the final equation. In fact, examination of the relative magnitude of the squared beta-weights for the predictors might be interpreted in such a manner as to suggest that $X_4$ accounted for the second highest proportion of criterion variance. Such an interpretation is obviously unrealistic from a theoretical viewpoint.

Some understanding of the role of $X_4$ in the equation can be gained from examining the residual correlation between this variable and the criterion with the effects of $X_5$ and $X_2$ removed by computation of first and second-order partial correlations. The correlation between $X_4$ and $C_4$ with the effects of $X_5$ and $X_2$ removed was $-0.05$. When considered in relation to the factor analysis, the magnitude of the residual correlation between $X_4$ and $C_4$ strongly suggests that partialing out the effects of $X_5$ and $X_2$ removed the variance in $X_4$ accounted for by the three creativity factors. However, the role of $X_4$ in the equation cannot be further determined with any reasonable degree of confidence since the nature of residual variance in $X_4$ is not evident from the present analysis.

Interpretation of the remainder of the regression equation is not presented. Several considerations entered into this decision. First, the beta-square interpretation for multiple regression equations can be justified only if it is used to generate or extend theoretical interpretations of the data. The remainder of the equation failed to make sense in
terms of either the factor analysis or the theory associated with the W-K creativity operations. Second, the behavior of the Pattern Meanings subtest \( X_4 \) was intimately related to the remainder of the equation. Interpretation of the variables selected after \( X_4 \) is potentially misleading since the nature of the residual variance in \( X_4 \) cannot be determined from the present analysis. Third, presentation of the regression equation as an empirical finding without interpretation serves to call attention to the need for cross-validation.

A satisfactory answer to the research question of concern in the present subsection can be formulated without full interpretation of the regression equation. The importance of the Line Meanings and Alternative Uses subtests in predicting Effectiveness of Expression ratings for clay products is clearly established in the case of the present sample. In specific, these two W-K subtests accounted for a substantial and highly significant proportion of criterion variance. The results of the hierarchical factor analysis clearly indicates that the c-factor and Vis. c-factor constitute the common denominator of performance on the Line Meanings and Alternative Uses subtests of the W-K and the criterion.

In addition, the magnitude of the beta-square for \( X_2 \) in the regression equation and the respective loadings of \( X_2 \) and \( C_4 \) on the Ver. c-factor suggest that this factor was involved in selection of the Alternative Uses subtest. The importance of the Ver. c-factor in the regression equation is,
however, necessarily limited since $C_4$ loads this factor to
a much smaller degree than the c-factor and Vis. c-factor.
Consequently, the Ver. c-factor figured less prominently
than the Vis. c-factor and c-factor in the selection pro-
cedure. Inspection of the hierarchical factor matrix suggests
that the Alternative Uses subtest was selected because it
constituted the best measure of verbal creativity. At the same
time, the Line Meanings subtest was selected because it pro-
vided the best estimate of both general and visual creativity.

An Overview of Clay Products: Comparison of the
factorial composition for the Originality ($C_3$) and Effective-
ness of Expression ($C_4$) dimensions provides a helpful frame-
work for examining similarities and differences in their re-
spective regression equations. Inspection of the hierarchical
factor matrix indicates remarkable similarity in the factorial
composition of $C_3$ and $C_4$.

The Vis. c-factor was the most important component
of performance on both the Originality and Effectiveness of
Expression criteria. Visual creativity accounted for 32 per
cent of the variance in the former and 33 per cent of the
variance in the latter. Thus, the tendency to produce a large
number of verbal associations to visual stimuli accounted for
the highest proportion of variance in both creativity criteria
derived from clay products. The prominence of the Line Mean-
ings subtest in the regression equations for both criteria can
be readily understood since it constituted the best measure of
visual creativity.

The c-factor was the second most important component of performance on both creativity criteria. General creativity accounted for 19 per cent of the variance in the Originality dimension and 24 per cent of the variance in the Effectiveness of Expression dimension. Thus, a generalized tendency to generate plentiful verbal associations to both verbal and visual stimuli constitutes an important component of performance on both creativity criteria derived from clay products. Selection of the Line Meanings subtest in both regression equations can again be understood since it also loaded stronger on the general creativity factor than any of the other predictor variables.

The Ver. c-factor constitutes a component of both creativity criteria derived from clay products. However, the verbal creativity component is clearly of minor importance when compared with either of the other creativity factors. This point is well illustrated by the fact that the Ver. c-factor accounted for only 7 per cent of the variance in C₃ and 12 per cent of that in C₄. The higher loading of C₄ on this factor suggests the possibility that verbal creativity may be a more important component of Effectiveness of Expression than Originality. In any case, it seems safe to conclude that the tendency to generate plentiful verbal associations to verbal stimuli constitutes a component of performance on both creativity criteria even though this component is considerably
less than that of the other two creativity factors.

The slight negative loadings of C_3 and C_4 on the g-factor suggest a negative relationship between general intelligence and performance on the two criteria derived from clay products. From this perspective, the negative regression weights for X_6 and X_10 in the equation for C_3 can be understood. It is also likely that the negative weight for X_6 and perhaps that of X_7 in the equation for C_4 can be understood in terms of this relationship.

Finally, both C_3 and C_4 showed a slight tendency to load Factor 7 which could not meaningfully be defined in terms of the variables which loaded it.

Thus, the high correlation between the Originality and Effectiveness of Expression criteria can be largely explained in terms of common variance on the three creativity factors. As previously indicated, the correlation between these two criteria was 0.889 thus indicating that 79 percent of the variance in C_3 and C_4 is held in common.

The parallel in factorial composition for C_3 and C_4 is also evident with regard to the g-factor and Factor 7 as well as the three creativity factors. This factorial similarity and high proportion of common variance for C_3 and C_4 raises doubt concerning interpretation of differences between the regression equations for these respective criteria. In other words, the results of the hierarchial factor analysis do not indicate any systematic differences be-
between C₃ and C₄ which might account for discrepancies between
the respective regression equations. Under such conditions
it seems that further attempts at either comparison or contrast
of the regression equations cannot be justified.

Crayon Drawings:

The original design for the study required detailed
analysis of the regression equations for all creativity cri-
teria. However, the instability of the creativity derived
from crayon drawings necessitated a modification in presenta-
tion of the data. The stepwise regression equations were de-
termined as originally specified but the results are tabulated in
Appendixes J-M but not interpreted in this section. The data
discussed in this section were obtained directly from the
31 x 31 predictor-criterion intercorrelation matrix rather
than the stepwise regression procedure.

This modification was indicated by conservative re-
search philosophy which dictates that weak and/or unstable
relationships should not be overinterpreted (Cronbach, 1968).
The instability of the criterion ratings for crayon drawings
poses a serious threat to the study in that poor control tends
to introduce potential confounding which cannot be eliminated
by even the most careful statistical analysis. In fact, Garrett
(1958) notes that poor experimental control constitutes the
most common source of spurious correlation.

Originality Ratings (C₁ and C₂): As previously in-
dicated, originality ratings were obtained for both sets of crayon drawings. Originality ratings for the first set of crayon drawings are referred to as C₁ while originality ratings for the second set are referred to as C₂. In operation terms, each pupil's score on a given criterion was obtained by summing the four judges' ratings to obtain a composite rating.

Examination of the correlations between C₁ and the respective predictor variables reveals that the Pattern Meanings subtest of the W-K is the most effective predictor. The correlation between this variable and the criterion was 0.243 which attained significance at the 0.05 level. However, the Pattern Meanings subtest explained only 6 per cent of the variance in C₁. This relationship provides weak evidence to support the validity of the W-K creativity operations.

The Line Meanings subtest of the W-K also showed a significant correlation with the criterion. In specific, this correlation was 0.237 which attained significance at the 0.05 level. This relationship was also very weak as it explains less than 6 per cent of the criterion variance. None of the remaining predictors showed a significant correlation with the criterion.

Examination of the correlations between C₂ and the respective predictors reveals that only the Instances subtest of the W-K correlates significantly with this criterion. The correlation between the criterion and the Instances subtest
was 0.235 which attained significance at the 0.05 level. Again, one sees a relationship which is significant but fails to explain a substantial proportion of criterion variance. None of the other predictors showed a significant correlation with the criterion.

In summary, the data presented in this subsection provides limited support for the validity of the W-K creativity operations. Yet, it seems reasonable to assume that the fault lies with the procedures employed in criterion development rather than the W-K creativity operations. That is, one must conclude that the present study failed to provide an adequate test for the validity of the W-K operations in predicting Originality ratings for crayon drawings.

**Effectiveness of Expression Ratings (C5 and C6):**

Inspection of the relationship between C5 and the respective predictor variables reveals that only the Block Design subtest of the WISC and the Instances subtest of the W-K correlated significantly with the criterion. The correlation between the Block Design subtest and the criterion was 0.254 which attained significance at the 0.05 level. This relationship explained approximately 6 per cent of the criterion variance. The correlation between the Instances subtest and the criterion was 0.247 which also attained significance at the 0.05 level and explained approximately 6 per cent of the criterion variance.

None of the predictors shows a significant correlation with C6. However, the Similarities subtest of the W-K
showed a correlation of 0.202 with the criterion which closely approaches significance. With the exception of this variable, none of the other predictors even approached significance.

These data provide little support for either establishing or refuting the validity of the W-K creativity operations. Here again, the instability of the criterion ratings resulted in confounding to the degree that one must acknowledge failure in providing an adequate validity test.
CHAPTER VI

CONCLUSIONS

The results of the present study clearly demonstrate the advantages of a hierarchical approach to creativity research based on Wherry's (1959) theory. Hierarchical factor analysis provides an opportunity for substantially increasing the amount of information which can be extracted from a given data matrix. That is, the investigator can not only determine empirical relationships but also go beyond to examine the nature of the relationship itself. Other methods of factor analysis can successfully identify the basic dimensions involved in a relationship but they have not proven successful in delineating the structural relations among these basic dimensions. Herein lies the decisive advantage of hierarchical factor analysis. It not only identifies the basic dimensions which account for a relationship but also provides information concerning the structural relations among the dimensions accounting for the relationship.

The advantages described above are clearly evident in the present study where information concerning the hierarchical factor structure of predictor and criterion variables provided a conceptual basis for interpreting the regression equations for clay products. Without this information one could have concluded only that the W-K creativity operation are relatively effective predictors of criterion performance and that the two criteria
obtained from clay products are highly related. Such information is not without value given the current confusion in creativity research. However, the main point is that the information obtained from the hierarchical analysis justifies a much broader set of conclusions which are discussed below.

First, the results of the hierarchical analysis served to identify the basic dimensions involved in both the W-K creativity operations and creativity criteria, to specify the structural relations among the dimensions of the creativity domain, and finally to relate the entire creativity domain to other aspects of cognition and background variables. According to Underwood (1957) such an approach is essential for the development of useful psychological theory as he states, "Scientific advance depends upon analysis and inevitably follows the initial identification of gross phenomena..." (p. 271). This information is directly relevant to the theoretical work of Mednick (1962), Wallach and Kogan (1965), Wallach and Wing (1969), and Wallach (1967, 1970).

For example, Mednick (1962) treats creativity as a unitary dimension of cognition throughout his discussion but digressed briefly at one point (p. 224) to note that it may be possible to differentiate between verbalizers and visualizers. He went on to surmise that differentiation along this dimension may well account for the fact that some individuals are creative in one area but not in another. The results of the hierarchical
analysis strongly confirm Mednick's hunch and suggest elaboration of this aspect of his theory. Furthermore, the importance of the Vis. c-factor suggests that effective prediction of criteria in visual arts may not be achieved unless measures of visual creativity are included.

Second, the results of the hierarchical analysis suggests that the task of the creativity researcher may well consist of determining the specific combination of factors which constitute creative performance on a specific criterion within a given area. More specifically, the researcher should be alerted to the possibility that he is likely to encounter not only a generalized disposition toward creativity (c-factor) but also visual creativity (Vis. c-factor) and verbal creativity (Ver. c-factor) along with area specific creativity, e.g. crayon drawings in the present study. A hierarchical orientation to creativity research has the advantage of focusing the investigator's attention on the empirical aspects of the area while providing a frame of reference by which his findings can be related to the work of other researchers.

Third, a hierarchical approach to creativity research encourages the investigator to assume what Underwood (1957, p. 271) describes as an analytic research orientation. This orientation is best characterized as a relentless analysis of gross phenomena into subphenomena through operational analysis. When basic dimensions are operationally defined they should be studied in detail to determine their generality. That is, the dimensions
should be carefully studied under different experimental conditions and across a wide range of samples. The various dimensions tend to accrue meaning in the course of such research. Meaningful creativity theory can, in the opinion of the writer, evolve only through such a deliberate process.

The congruence between Underwood's (1957) analytic orientation and Wherry's (1959) hierarchical model is readily evident. Both approaches stress the reduction of gross phenomena to their basic dimension, establishment of structural relationships among these dimensions, and study of these structural relationships under varying conditions. The two approaches are complementary as Underwood focuses more specifically on the operational aspects of research while Wherry emphasizes identification of the structural relationships among the basic dimensions.

Several implications for criterion development can be readily deduced from the results of the present study when considered in the context of the above discussion. For example, the present attempt at criterion development was successful only to the extent that a gross indication of creative performance was obtained for clay products. An analytic approach dictates that a wide variety of criterion ratings should be secured for this or any other area of creative endeavor. Determination of structural relationships among these criteria should provide the information necessary for understanding the nature of creative performance. The stability of both the products and the
judges' ratings over time should also be carefully determined.

After the above procedure is accomplished for a given area further validation should be undertaken across a series of homogeneous samples. A wide range of control and intelligence variables should also be included for each sample along with the W-K creativity operations. The intelligence assessors would function as markers to define the g-factor and subgeneral intelligence factors. The control variables would function to identify the potential influence of background and situational variables on criterion performance. Data regarding the past history and current status should serve to suggest research hypothesis regarding the antecedents of creative performance.

Implementation of such a research program might best be accomplished by studying a series of homogeneous subsamples at a given age level provided these subsamples were chosen to differ systematically with regard to relevant control variables. A sequence of such studies across various age groups should serve to dispel much of the confusion currently evident among creativity researchers. Furthermore, such a strategy should provide the information necessary for the design of meaningful longitudinal research.

Thus, it seems reasonable to conclude that Wherry's (1959) hierarchical theory of cognition provides the most promising strategy currently available for creativity research. In this regard, the results of the present study are in agreement with those of Haynes (1970) who reached essentially the same
conclusion for the area of intelligence after application of hierarchial factor analysis to a battery of Guilford's tests.
Appendix A: Distribution of Intelligence Quotients for the Sample

<table>
<thead>
<tr>
<th>Score interval</th>
<th>Frequency</th>
<th>Cumulative frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>125-129</td>
<td>2</td>
<td>73</td>
</tr>
<tr>
<td>120-124</td>
<td>2</td>
<td>71</td>
</tr>
<tr>
<td>115-119</td>
<td>5</td>
<td>69</td>
</tr>
<tr>
<td>110-114</td>
<td>2</td>
<td>64</td>
</tr>
<tr>
<td>105-109</td>
<td>8</td>
<td>62</td>
</tr>
<tr>
<td>100-104</td>
<td>19</td>
<td>54</td>
</tr>
<tr>
<td>95-99</td>
<td>16</td>
<td>35</td>
</tr>
<tr>
<td>90-94</td>
<td>9</td>
<td>19</td>
</tr>
<tr>
<td>85-89</td>
<td>5</td>
<td>10</td>
</tr>
<tr>
<td>80-84</td>
<td>4</td>
<td>5</td>
</tr>
<tr>
<td>75-79</td>
<td>1</td>
<td>1</td>
</tr>
</tbody>
</table>
Appendix B: Distribution of Birth Order Variable for the Sample

<table>
<thead>
<tr>
<th>Birth position</th>
<th>Frequency</th>
<th>Cumulative frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>8</td>
<td>2</td>
<td>73</td>
</tr>
<tr>
<td>7</td>
<td>1</td>
<td>71</td>
</tr>
<tr>
<td>6</td>
<td>2</td>
<td>70</td>
</tr>
<tr>
<td>5</td>
<td>4</td>
<td>68</td>
</tr>
<tr>
<td>4</td>
<td>17</td>
<td>64</td>
</tr>
<tr>
<td>3</td>
<td>15</td>
<td>47</td>
</tr>
<tr>
<td>2</td>
<td>20</td>
<td>32</td>
</tr>
<tr>
<td>1</td>
<td>12</td>
<td>12</td>
</tr>
</tbody>
</table>
### Appendix C: Distribution ofChronological Age for
the Sample

<table>
<thead>
<tr>
<th>Interval</th>
<th>Frequency</th>
<th>Cumulative frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>140-144</td>
<td>1</td>
<td>73</td>
</tr>
<tr>
<td>135-139</td>
<td>2</td>
<td>72</td>
</tr>
<tr>
<td>130-134</td>
<td>5</td>
<td>70</td>
</tr>
<tr>
<td>125-129</td>
<td>7</td>
<td>65</td>
</tr>
<tr>
<td>120-124</td>
<td>12</td>
<td>58</td>
</tr>
<tr>
<td>115-119</td>
<td>13</td>
<td>46</td>
</tr>
<tr>
<td>110-114</td>
<td>17</td>
<td>33</td>
</tr>
<tr>
<td>105-109</td>
<td>5</td>
<td>16</td>
</tr>
<tr>
<td>100-104</td>
<td>11</td>
<td>11</td>
</tr>
</tbody>
</table>
Appendix D: Rating Procedure for Art Products

Part I: Briefing Procedure for Judges:

The first step in the rating procedure consisted of an informal meeting at which time the investigator and art consultant explained the procedure to be followed in rating the art products. Each judge was presented with copies of the following: 1) definitions of the Originality and Effectiveness of Expression dimensions, 2) the forms to be used in recording their ratings, and 3) the instructions for the actual rating procedure to be followed.

At this time the judges were told that they were to rate three sets of creative products on each of the two dimensions described to them. They were also told that their ratings for each set of products was to be presented to the investigator when their ratings were complete and that they would then be conducted to the room containing the next set of products.

The art consultant and examiner answered questions posed by the judges at the end of the initial presentation. Several questions were raised concerning the meanings of the definitions of Originality and Effectiveness of Expression. The art consultant discussed the questions with the judges in an attempt to assure that they approached the task with the same interpretation of the definitions.
Part II: Definitions of Criterion Dimensions:

1. **Originality:** Your ratings of this dimension should reflect the extent to which you sense originality, novelty, and/or uniqueness in the ideas underlying the product.

2. **Effectiveness of Expression:** Your ratings of this dimension should reflect the extent to which you sense aesthetic quality, exquisite quality in communication, and/or an impression of beauty in the use of elements or media.

Part III: Directions for Ratings

You are to rate each product on the specified dimension from 9 (highest through 5 (average) to 1 (inferior). You have a copy of the form to be used in recording your ratings. Arrange the products into nine groups corresponding to the nine categories on the record form. That is you should arrange the products into nine groups which might be described as Superior, Very Good, Good, High Average, Average, Low Average, Poor, Very Poor, and Inferior. Try to arrange your groups in such a manner that you have about the same number of products in each category. Since there are 73 products in each set and nine categories you should have between 8 and 10 products in each category.

Your ratings are to be based on the quality of a given product in relation to that of the other products in the set. Therefore, it is very important for you to take some time to look over all of the products in a set before making your ratings. Take as much time as you like to complete your ratings.

When you have completed your ratings for a given set record the product number in the appropriate column of the record form, sign the form, and bring it to me. I will take you
to the room where the next set of products is located.
### Appendix F: Distribution of z-Scores for the Five Creativity Subtests

<table>
<thead>
<tr>
<th>Score interval</th>
<th>0.00 to 4.49</th>
<th>3.50 to 3.99</th>
<th>3.00 to 3.49</th>
<th>2.50 to 2.99</th>
<th>2.00 to 2.49</th>
<th>1.50 to 1.99</th>
<th>1.00 to 1.49</th>
<th>0.50 to 0.99</th>
<th>0.00 to 0.49</th>
<th>-0.01 to -0.50</th>
<th>-0.51 to -1.00</th>
<th>-1.01 to -1.50</th>
<th>-1.51 to -2.00</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>15</td>
<td>12</td>
<td>11</td>
<td>19</td>
</tr>
<tr>
<td></td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>25</td>
<td>16</td>
<td>19</td>
<td>21</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>4</td>
<td>8</td>
<td>10</td>
<td>4</td>
</tr>
<tr>
<td></td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Appendix H: Regression Equation for C3 Based on Conservative Decision Rule

<table>
<thead>
<tr>
<th>Variable</th>
<th>Beta-Weight</th>
<th>Bee-Weight</th>
<th>t for Weight</th>
</tr>
</thead>
<tbody>
<tr>
<td>5</td>
<td>0.556</td>
<td>0.234</td>
<td>5.740</td>
</tr>
<tr>
<td>2</td>
<td>0.211</td>
<td>0.126</td>
<td>2.180</td>
</tr>
</tbody>
</table>

A^2 Weight: 7.182
Appendix I: Regression Equation for C4 Based on Conservative Decision Rule

<table>
<thead>
<tr>
<th>Variable</th>
<th>Beta-Weight</th>
<th>Bee-Weight</th>
<th>t for Weight</th>
</tr>
</thead>
<tbody>
<tr>
<td>5</td>
<td>0.548</td>
<td>0.227</td>
<td>6.257</td>
</tr>
<tr>
<td>2</td>
<td>0.303</td>
<td>0.178</td>
<td>3.456</td>
</tr>
<tr>
<td>7</td>
<td>-0.181</td>
<td>-0.730</td>
<td>-2.257</td>
</tr>
</tbody>
</table>

A-Weight: 12.969
**Appendix J: Full Stepwise Regression Procedure for Cl**

<table>
<thead>
<tr>
<th>Variable</th>
<th>R</th>
<th>$\bar{R}$</th>
<th>F</th>
<th>df</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>4</td>
<td>.243</td>
<td>.214</td>
<td>4.470</td>
<td>1,71</td>
<td>.05</td>
</tr>
<tr>
<td>18</td>
<td>.303</td>
<td>.256</td>
<td>2.489</td>
<td>2,70</td>
<td>.01</td>
</tr>
<tr>
<td>10</td>
<td>.348</td>
<td>.288</td>
<td>2.337</td>
<td>3,69</td>
<td>.10</td>
</tr>
<tr>
<td>7</td>
<td>.369</td>
<td>.292</td>
<td>1.183</td>
<td>4,68</td>
<td>.50</td>
</tr>
<tr>
<td>17</td>
<td>.403</td>
<td>.314</td>
<td>2.101</td>
<td>5,67</td>
<td>.10</td>
</tr>
<tr>
<td>19</td>
<td>.422</td>
<td>.321</td>
<td>1.213</td>
<td>6,66</td>
<td>.50</td>
</tr>
</tbody>
</table>
### Appendix K: Regression Equation for C1 Obtained from Full Stepwise Regression Procedure

<table>
<thead>
<tr>
<th>Variable</th>
<th>Beta-Weight</th>
<th>Bee-Weight</th>
<th>t for weight</th>
</tr>
</thead>
<tbody>
<tr>
<td>4</td>
<td>0.211</td>
<td>0.089</td>
<td>1.784</td>
</tr>
<tr>
<td>18</td>
<td>0.214</td>
<td>0.153</td>
<td>1.687</td>
</tr>
<tr>
<td>10</td>
<td>0.251</td>
<td>0.728</td>
<td>1.848</td>
</tr>
<tr>
<td>7</td>
<td>-0.211</td>
<td>-0.644</td>
<td>-1.588</td>
</tr>
<tr>
<td>17</td>
<td>0.174</td>
<td>0.476</td>
<td>1.411</td>
</tr>
<tr>
<td>19</td>
<td>0.125</td>
<td>1.793</td>
<td>1.101</td>
</tr>
</tbody>
</table>

A-Weight: 10.013
### Appendix L: Full Stepwise Regression Procedure for C2

<table>
<thead>
<tr>
<th>Variable</th>
<th>R</th>
<th>(\bar{R})</th>
<th>F</th>
<th>df</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>.235</td>
<td>.235</td>
<td>4.152</td>
<td>1,71</td>
<td>.05</td>
</tr>
<tr>
<td>6</td>
<td>.296</td>
<td>.249</td>
<td>2.506</td>
<td>2,70</td>
<td>.10</td>
</tr>
<tr>
<td>10</td>
<td>.341</td>
<td>.279</td>
<td>2.222</td>
<td>3,69</td>
<td>.10</td>
</tr>
<tr>
<td>9</td>
<td>.370</td>
<td>.293</td>
<td>1.169</td>
<td>4,68</td>
<td>.50</td>
</tr>
<tr>
<td>8</td>
<td>.394</td>
<td>.304</td>
<td>1.463</td>
<td>5,67</td>
<td>.25</td>
</tr>
<tr>
<td>14</td>
<td>.420</td>
<td>.319</td>
<td>1.711</td>
<td>6,66</td>
<td>.25</td>
</tr>
<tr>
<td>23</td>
<td>.449</td>
<td>.341</td>
<td>2.054</td>
<td>7,65</td>
<td>.10</td>
</tr>
<tr>
<td>12</td>
<td>.465</td>
<td>.344</td>
<td>1.151</td>
<td>8,64</td>
<td>.25</td>
</tr>
<tr>
<td>15</td>
<td>.482</td>
<td>.351</td>
<td>1.359</td>
<td>9,63</td>
<td>.25</td>
</tr>
<tr>
<td>20</td>
<td>.502</td>
<td>.363</td>
<td>1.643</td>
<td>10,62</td>
<td>.25</td>
</tr>
<tr>
<td>19</td>
<td>.519</td>
<td>.371</td>
<td>1.400</td>
<td>11,61</td>
<td>.25</td>
</tr>
</tbody>
</table>
Appendix M: Regression Equation for C2 Obtained from Full Stepwise Regression Procedure

<table>
<thead>
<tr>
<th>Variable</th>
<th>Beta-Weight</th>
<th>Bee-Weight</th>
<th>t for Weight</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.284</td>
<td>0.069</td>
<td>2.435</td>
</tr>
<tr>
<td>6</td>
<td>-0.290</td>
<td>-0.974</td>
<td>-2.075</td>
</tr>
<tr>
<td>10</td>
<td>0.107</td>
<td>0.346</td>
<td>0.653</td>
</tr>
<tr>
<td>9</td>
<td>-0.296</td>
<td>-0.986</td>
<td>-2.139</td>
</tr>
<tr>
<td>8</td>
<td>0.362</td>
<td>0.987</td>
<td>2.341</td>
</tr>
<tr>
<td>14</td>
<td>-0.221</td>
<td>-0.703</td>
<td>-1.598</td>
</tr>
<tr>
<td>23</td>
<td>0.232</td>
<td>0.718</td>
<td>1.866</td>
</tr>
<tr>
<td>12</td>
<td>0.176</td>
<td>0.440</td>
<td>1.303</td>
</tr>
<tr>
<td>15</td>
<td>-0.200</td>
<td>-0.600</td>
<td>-1.510</td>
</tr>
<tr>
<td>20</td>
<td>-0.161</td>
<td>-2.541</td>
<td>-1.409</td>
</tr>
<tr>
<td>19</td>
<td>0.134</td>
<td>2.133</td>
<td>1.183</td>
</tr>
</tbody>
</table>

A-Weight: 23,324
Appendix N: Full Stepwise Regression Procedure for C5

<table>
<thead>
<tr>
<th>Variable</th>
<th>R</th>
<th>R</th>
<th>F</th>
<th>df</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>14</td>
<td>.254</td>
<td>.227</td>
<td>4.908</td>
<td>1.71</td>
<td>.05</td>
</tr>
<tr>
<td>1</td>
<td>.358</td>
<td>.321</td>
<td>5.082</td>
<td>2.70</td>
<td>.01</td>
</tr>
<tr>
<td>22</td>
<td>.402</td>
<td>.353</td>
<td>2.741</td>
<td>3.69</td>
<td>.05</td>
</tr>
<tr>
<td>13</td>
<td>.453</td>
<td>.398</td>
<td>3.743</td>
<td>4.68</td>
<td>.01</td>
</tr>
<tr>
<td>19</td>
<td>.483</td>
<td>.419</td>
<td>2.456</td>
<td>5.67</td>
<td>.05</td>
</tr>
<tr>
<td>18</td>
<td>.497</td>
<td>.423</td>
<td>1.226</td>
<td>6.66</td>
<td>.50</td>
</tr>
<tr>
<td>20</td>
<td>.515</td>
<td>.431</td>
<td>1.559</td>
<td>7.65</td>
<td>.25</td>
</tr>
<tr>
<td>15</td>
<td>.526</td>
<td>.431</td>
<td>1.009</td>
<td>8.64</td>
<td>.50</td>
</tr>
</tbody>
</table>
Appendix O: Regression Equation for C5 Obtained from Full Stepwise Regression Procedure

<table>
<thead>
<tr>
<th>Variable</th>
<th>Beta-Weight</th>
<th>Bee-Weight</th>
<th>t for Weight</th>
</tr>
</thead>
<tbody>
<tr>
<td>14</td>
<td>0.308</td>
<td>0.797</td>
<td>2.635</td>
</tr>
<tr>
<td>1</td>
<td>0.176</td>
<td>0.035</td>
<td>1.511</td>
</tr>
<tr>
<td>22</td>
<td>0.310</td>
<td>4.600</td>
<td>2.638</td>
</tr>
<tr>
<td>13</td>
<td>-0.227</td>
<td>-0.474</td>
<td>-1.891</td>
</tr>
<tr>
<td>19</td>
<td>-0.187</td>
<td>-2.409</td>
<td>-1.713</td>
</tr>
<tr>
<td>18</td>
<td>0.160</td>
<td>0.104</td>
<td>1.367</td>
</tr>
<tr>
<td>20</td>
<td>-0.118</td>
<td>-1.515</td>
<td>-1.060</td>
</tr>
<tr>
<td>15</td>
<td>0.124</td>
<td>0.302</td>
<td>1.004</td>
</tr>
</tbody>
</table>

A-Weight: -2.100
Appendix P: Full Stepwise Regression Procedure for C6

<table>
<thead>
<tr>
<th>Variable</th>
<th>R</th>
<th>R</th>
<th>F</th>
<th>df</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>3</td>
<td>.202</td>
<td>.166</td>
<td>3.304</td>
<td>1.71</td>
<td>.10</td>
</tr>
<tr>
<td>2</td>
<td>.255</td>
<td>.196</td>
<td>1.811</td>
<td>2.70</td>
<td>.25</td>
</tr>
<tr>
<td>13</td>
<td>.297</td>
<td>.221</td>
<td>1.752</td>
<td>3.69</td>
<td>.25</td>
</tr>
<tr>
<td>15</td>
<td>.335</td>
<td>.244</td>
<td>1.809</td>
<td>4.68</td>
<td>.25</td>
</tr>
<tr>
<td>19</td>
<td>.358</td>
<td>.251</td>
<td>1.236</td>
<td>5.67</td>
<td>.50</td>
</tr>
<tr>
<td>25</td>
<td>.378</td>
<td>.255</td>
<td>1.155</td>
<td>6.66</td>
<td>.50</td>
</tr>
<tr>
<td>9</td>
<td>.395</td>
<td>.255</td>
<td>1.005</td>
<td>7.65</td>
<td>.50</td>
</tr>
<tr>
<td>7</td>
<td>.419</td>
<td>.269</td>
<td>1.485</td>
<td>8.64</td>
<td>.25</td>
</tr>
</tbody>
</table>
Appendix Q: Regression Equation for C6 Obtained from Full Stepwise Regression Procedure

<table>
<thead>
<tr>
<th>Variable</th>
<th>Beta-Weight</th>
<th>Bee-Weight</th>
<th>t for Weight</th>
</tr>
</thead>
<tbody>
<tr>
<td>3</td>
<td>9.277</td>
<td>0.128</td>
<td>1.429</td>
</tr>
<tr>
<td>2</td>
<td>-0.345</td>
<td>-0.155</td>
<td>-1.795</td>
</tr>
<tr>
<td>13</td>
<td>-0.170</td>
<td>-0.398</td>
<td>-1.436</td>
</tr>
<tr>
<td>15</td>
<td>0.133</td>
<td>0.362</td>
<td>0.982</td>
</tr>
<tr>
<td>19</td>
<td>-0.202</td>
<td>-2.919</td>
<td>-1.621</td>
</tr>
<tr>
<td>25</td>
<td>-0.240</td>
<td>-1.062</td>
<td>-1.533</td>
</tr>
<tr>
<td>9</td>
<td>-0.155</td>
<td>-0.468</td>
<td>-1.245</td>
</tr>
<tr>
<td>7</td>
<td>0.156</td>
<td>0.480</td>
<td>1.219</td>
</tr>
</tbody>
</table>

A-Weight: 25.379
REFERENCES


Harmon, L. The development of a criterion of scientific competence. In C. Taylor & F. Barron (Eds.), *Scientific creativity*:


Mednick, S. The associative basis of the creative process.  

Mischel, W., & Mitzner, R. Preference for delayed reward as a 
function of age, intelligence, and length of delay interval.  


Pankove, E. _The relationship between creativity and risk taking in 
fifth-grade children_. (Doctoral dissertation, Rutgers University) 

Pankove, E., & Kogan, N. Creative ability and risk taking in 
elementary school children. _Journal of Personality_, 1968,  
36, 420-439.


Roe, A. A psychological study of eminent psychologists and anthropologists and a comparison with biological and physical scientists.  
_Psychological Monographs_, 1953, 67, No. 2.

Schachter, S. Birth order, eminence and higher education. _American 


Skager, R., Klein, S., & Schultz, C. The prediction of academic and artistic achievement at a school of design. _Journal of Education- 
al Measurements_, 1967, 4, 105-117.


Spotts, J., & Mackler, B. Relationships of field-dependent and 

Stein, M. Creativity. In E. Borgatta & W. Lambert (Eds.),  
_Handbook of personality theory and research_. Chicago: Rand
McNally, 1968.


Thorndike, R. The measurement of creativity. Teachers College Record, 1963, 64, 422-424.


Wherry, R. J. WETSL program. In R. J. Wherry (Ed.), *Psychology department computer programs*. Columbus, Ohio: Department of Psychology, The Ohio State University, 1969.

Wherry, R. J. CHIRAT program. In R. J. Wherry (Ed.), *Psychology department computer programs*. Columbus, Ohio: Department of Psychology, The Ohio State University, 1970.
Wherry, R. J. and Wherry, R. J., Jr. WHEWR program. In R. J. Wherry (Ed.), *Psychology department computer programs*. Columbus, Ohio: Department of Psychology, The Ohio State University, 1968.


