# THE EFFECTS OF LACK OF FORMAL SCHOOL EXPERIENCE ON PERFORMANCE ON TESTS OF CREATIVE THINKING

THESIS FOR THE DEGREE OF PH. D.

MICHIGAN STATE UNIVERSITY

MILTON C. HILLERY

1969

This is to certify that the

thesis entitled

THE EFFECTS
OF
LACK OF FORMAL SCHOOL EXPERIENCE ON
PERFORMANCE ON TESTS OF CREATIVE THINKING

presented by

MILTON C. HILLERY

has been accepted towards fulfillment
of the requirements for

Ph.D. degree in Education

Lee S. Shulman
Major professor

Date May 13, 1969

ABSTRACT

THE EFFECTS OF LACK OF FORMAL SCHOOL EXPERIENCE ON PERFORMANCE ON TESTS OF CREATIVE THINKING

by Milton C. Hillery

This study attempted to test a number of hypotheses relating to the differential performance of two major groups of elementary school children on tests of creative thinking.

The rationale developed from the related literature focused upon formal school experience as a suppressant of creative development.

The two major groups tested consisted of children from Freedom Schools in Prince Edward County, Virginia and subjects taken from the elementary schools of Jackson, Michigan. At the time of testing the public schools in Virginia had been closed for five years.

The Virginia sample group consisted of four sub-groups based on age and grade level: (1) Grade 1 - children who essentially had not missed any formal school experience (kindergarten excepted); (2) Grade 4 - children who had missed three years and were in their first year of school; (3) Grade 6 - children who had been in school one year, out four years, and in their second year of school; and (4) Grade 8 - students who had been in school three years, out four years, and in their fourth year of school.

The Michigan sample group was also subdivided into four groups based on age and grade-level identical to those groups in Virginia. However, the Michigan children had had the normal number of years of formal school experience implied by their grade level status.

The criterion measures for creative thinking used were the four tasks of the Minnesota Test of Creative Thinking, Abbreviated Form VII, developed by E. P. Torrance of the University of Minnesota.

Two of the tasks are categorized as nonverbal ("Incomplete Figures" and "Circles") while the other two are verbal tasks ("Product Improvement" and "Unusual Uses").

Five basic scores were generated by the scoring process: (1) fluency, number of responses; (2) flexibility, number of different ideas represented by the responses; (3) originality, the relative novelty or uniqueness of the responses; (4) elaboration, the number of details or embellishments added to the original response, and (5) total score, the sum of the above four scale scores. The analysis was done on these five scores, computed separately for the verbal and nonverbal tasks.

The two sample groups were assumed to be comparable except for differences in amount of formal schooling. The purpose of the analysis was to determine whether there were detectable differences between the two groups on the criterion measures. An analysis of variance design was used to achieve these objectives. The design is that of a five factor repeated measure design with two of the factors comprising the repeated measures. The independent factors of county (Virginia, Michigan), sex, and age are completely crossed as are the repeated measures of scale and form (verbal and nonverbal). This results in a 2x2x4x4x2 design.

The major hypotheses dealt with performance differences between the Virginia and Michigan samples collectively and at each age or grade level. Additional hypotheses dealt with related factors such as age or sex, that might differentially affect performance.

The study also contained several secondary analyses and discussion; among these are an analysis of the relationship between I.Q. and Creativity test scores.

To summarize the major conclusions it can be stated that those children who had normal school experience, in terms of numbers of years of continuous schooling, did better on tests of creativity than did those children whose schooling had been delayed or interrupted for varying amounts of time. The differences between the two major populations were not large, and not even significant for the overall population. Beyond this fact, the results of the secondary analyses and other observations indicate that much of the differences in performance are a consequence of secondary or indirect effects of schooling, e.g., I.Q., vocabulary, pencil manipulation, etc.

The one general conclusion that would seem most warranted from the results of this study would be that the evidence indicates that in the given situation formal schooling has little, if any, direct positive effect on creative thinking, as measured by the tests utilized.

# THE EFFECTS OF LACK OF FORMAL SCHOOL EXPERIENCE ON PERFORMANCE ON TESTS OF CREATIVE THINKING

by

Milton C. Hillery

A THESIS

Submitted to
Michigan State University
in partial fulfillment of the requirements
for the degree of

DOCTOR OF PHILOSOPHY

Department of Counseling, Personnel Services and
Educational Psychology - College of Education

1969

## ACKNOWLEDGMENTS

For a study such as this, involving cooperation from several school systems, a personal acknowledgment of all those involved would be a very lengthy task, although a justifiable one. However a general acknowledgment must suffice for those in the administration and classrooms of the Freedom Schools in Prince Edward County, Virginia, and the Jackson, Michigan Public Schools. Their approval, cooperation and assistance made this study possible.

More specifically, the author is indebted to Dr. Robert Green and Dr. Louis Hofmann, then of Michigan State University, director and co-director of the research project "The Educational Status of Children in a District Without Public Schools," United States Office of Education, Cooperative Research Project, #2321, Department of Health, Education and Welfare. The financing of the project to be presented here was provided primarily from funds from this grant directed by Drs. Green and Hofmann; in addition, their theoretical and methodological assistance was major, as was their help in coordinating the tasks in Virginia.

Dr. Irving E. Sigel, of Merrill-Palmer, Detroit, Michigan and Dr. Ed Gordon, then of Yeshiva University, New York City, also made significant contributions to the design of the study.

Mrs. Lou Ingalls, a fellow graduate student at Michigan State University provided vital assistance in not only administering the group tests in Jackson, Michigan but also coordinating the work of the other two examiners; a special note of appreciation to those two undergraduate students at Michigan State, Miss Linda McClelland and Miss Mary McGee.

The contributions of a committee chairman to a doctoral dissertation need no elaboration, other than the fact that this project was conducted under the leadership of two chairmen. At the beginning of the project Dr. Lorraine Shepard was the author's advisor. At the time that Dr. Shepard left the campus, Dr. Lee Shulman, who had advised the author from the earliest development of the project, assumed the chairman's role. His guidance and general assistance have primarily shaped the final product. The author's indebtedness to both Dr. Shepard and Dr. Shulman far exceed this study. Each in their own way have been the primary factors in the professional growth of the author.

The laborious and frustrating job of preparing this manuscript from the earliest handwritten draft to its final form has been the undertaking of Mrs. Irene Sylvester.

TABLE OF CONTENTS


| | |
|---|---|
| Acknowledgments | ii |
| Table of Contents | iv |
| List of Tables | v |
| List of Figures | vii |
| List of Appendices | ix |
| Chapter 1 - Introduction | 1 |
| Chapter 2 - Review of the Literature | 4 |
| Chapter 3 - Procedures | 25 |
| Chapter 4 - Primary Analysis of the Data | 56 |
| Chapter 5 - Secondary Analysis of the Data | 84 |
| Chapter 6 - Discussion | 105 |
| Bibliography | 130 |
| Appendix A - Results of the Reliability Study | 134 |
| Appendix B - Procedure Developed to Explain the Three-Way Interaction of "County x Sex x Age" | 138 |
| Appendix C - Analysis of Creativity Task Scores | 144 |
| Appendix D - Minnesota Test of Creative Thinking, Abbreviated Form VII | 150 |

LIST OF TABLES


| Table | | Page |
|---|---|---|
| 1 | Format used for recording individual scores and derivation of scores in statistical analysis | 52 |
| 2 | Analysis of variance model employed | 54 |
| 3 | Means and standard deviations for fifteen creative performance scores for the 16 sample groups (County x Sex x Age) | 59-60 |
| 4 | Results of analysis of variance (ANOVA Table) | 61-62 |
| 5 | Analysis of differences of pairs of means for the eight groups (Age x County) using the Neuman-Keuls procedure | 65-65 |
| 6 | Analysis of differences of pairs of mean scores for subjects grouped by age - using the Neuman-Keuls procedure | 66 |
| 7 | Neuman-Keuls test for significance of difference in verbal and nonverbal scores by county | 70 |
| 8 | Analysis of differences of pairs of means for performance scores on the four scales, using the Neuman-Keuls procedure | 75 |
| 9 | Comparative scores for the three counties | 86 |
| 10 | Comparative I.Q. scores for Michigan and Virginia students by age levels | 90 |
| 11 | Rank order correlations between I.Q. and certain performance scores on the Minnesota Test of Creativity | 92 |
| 12 | Inter-task correlations total sample | 96 |
| 13 | Inter-task correlations for Virginia students | 97 |
| 14 | Inter-task correlations for Michigan students | 97 |
| 15 | Correlations of individual creativity task scores with total creativity scores for Virginia and Michigan children | 98 |
| 16 | Correlations of I.Q. with each of the four Creativity Task Scores | 98 |

| Table | | Page |
|---|---|---|
| 17 | Intercorrelations between Total Scale Scores | 100 |
| 18 | Intercorrelations between Total Scale Scores for Virginia children | 101 |
| 19 | Intercorrelations between Total Scale Scores for Michigan children | 101 |
| 20 | Correlations for Michigan and Virginia children between Total Scale Scores and Total Creativity Scores | 102 |
| 21 | Correlations for Michigan and Virginia children between Scale Scores and I.Q. | 102 |
| 22 | Test-retest relaibilities obtained on nonverbal tests of Minnesota Test of Creative Thinking, Abbreviated Form VII | 135-136 |
| 23 | Expected means by age by sex | 139 |
| 24 | Deviation of observed means from expected means with data grouped by age, sex and county | 140 |
| 25 | Table for derivation of expected mean values for data grouped by county by sex by age | 141 |
| 26 | Deviation of observed means from expected means for data grouped by county by age by sex | 142 |

LIST OF FIGURES


| Figure | | Page |
|---|---|---|
| 1 | Form performance scores by age | 68 |
| 2 | Verbal and nonverbal performance scores for Virginia and Michigan students by age level | 71 |
| 3 | Performance scores for Virginia and Michigan students, illustrated by sex and across age groups | 73 |
| 4 | Total scale scores by county | 77 |
| 5 | Mean creativity scores by age and group | 78 |
| 6 | Mean age group scores on the four creativity scores | 78 |
| 7 | Scale performance for 6 year old children | 79 |
| 8 | Scale performance for 9 year old children | 79 |
| 9 | Scale performance for 11 year old children | 79 |
| 10 | Scale performance for 13 year old children | 79 |
| 11 | Interaction effect of Scale x Form | 81 |
| 12 | Interaction effect of Scales x Form x County | 82 |
| 13 | Mean scale performance scores for first grade students on the Figure Completion task | 87 |
| 14 | Mean scale performance scores for fourth grade students on the Figure Completion task | 87 |
| 15 | Mean scale performance scores for sixth grade students on the Figure Completion task | 87 |
| 16 | Mean total scores for age levels 6, 9, & 11 on the Figure Completion task | 87 |
| 17 | Sex differences in obtained coefficients of reliability for total scale scores & total score of the nonverbal tasks of the Minnesota Tests of Creative Thinking, Abbreviated Form VII | 137 |
| 18 | Age differences in obtained coefficients of reliability for total scale scores and total score of the nonverbal tasks of the Minnesota Tests of Creative Thinking, Abbreviated Form VII | 137 |

Figure | Page

19 - Age group mean scores on Incomplete Figures Task for four subgroups (County x Sex) 145

20 - Age group mean scores on Circles Task for four subgroups (County x Sex) 145

21 - Age group mean scores on Product Improvement task for four subgroups (County x Sex) 145

22 - Age group mean scores on Uses Task for four subgroups (County x Sex) 145

23 - Task performance differentiated by county 147

24 - Scores on creativity tasks for each of four age groups 148

25 - Scores on creativity tasks by sex 149

## LIST OF APPENDICES


| Appendix | Page |
|---|---|
| A - Results of Reliability Study | 134 |
| B - Procedure Developed to Explain the Three-Way Interaction of "County x Sex x Age" | 138 |
| C - Analysis of Creativity Task Scores | 144 |
| D - Minnesota Test of Creative Thinking, Abbreviated Form VII | 150 |

# Chapter 1 - Introduction

## Statement of Problem

The purpose of this study is to assess the effect of lack of formal schooling upon performance on certain cognitive tasks generally labeled "creative."

The last five to ten years have witnessed an increased concern for the phenomenon labeled as "creativity." This increase has been demonstrated in part by the increase in numbers of publications on the subject in the professional journals of psychology and education, and also in the heightened concern of educators as to the role creative thinking, its development, or its maintenance should play in curricular decisions. The literature has failed to provide a clear cut definition of just what "creativity" is, or how such a trait can be measured. However, this same literature has been quite consistent in depicting the general process of formal schooling, characteristic of American society, as a suppressant of both creative development and creative productivity. (Anderson, 1961; Torrance, 1952a, 1963b, 1963c; Schachtel, 1959; Getzels and Jackson, 1960; Rugg, 1963; Drews, 1961; Wolfe, 1960; Gram, 1963; Barron, 1955; and Getzels, 1964.)

These accusations have spurred many educators to examine the relative effects of different situational variables on various manifestations of creative behavior. To do this the examiners have generally varied significant aspects of the educative process for one group and held the process constant for another, and then compared the two groups on the basis of some form of assessment of creative gain. A more thorough test of the position that formal school experiences are a deterrent or suppressant of creative development and productivity could be achieved by comparing students who have not been exposed to formal schooling for a significant period of time with those who have had the normal exposure. Such a comparison is the purpose of this study.

With the American vision of universal education having become more a reality than a dream, the opportunity to locate and assess a group not exposed to formal schooling is virtually non-existent. Such individuals are limited to a relatively isolated few in any given school area and are generally characterized by highly atypical traits which would confound any results obtained. One of the few opportunities to make such a comparison was presented as a consequence of the events in Prince Edward County, Virginia. A majority of the Negro students of this county did not have an opportunity for formal educational experiences as a result of the closing of the public schools. This population then provides a sample for comparison with a group whose involvement in formal schooling has been continuous and uninterrupted.

## Objectives

The specific objectives of this study can best be expressed by an indication of the nature of the hypotheses to be tested. To facilitate this preview of the hypotheses it is necessary to specify briefly the characteristics of the samples to be tested. Each sample group is

composed of four sub-groups distinguished by age; 6, 9, 11 and 13 year olds. (Or grade levels 1, 4, 6 and 8.) The Prince Edward County sample contains children who, according to their age level have missed three or four years of formal school experience, with the exception of the six year olds. The second sample, taken from the school population of Jackson, Michigan, is composed of children who have had the normal number of years of formal school experience for their age.

The hypotheses to be tested predict that, in general, the Prince Edward County samples will obtain significantly higher scores on the creative tasks than the Michigan sample.

## Chapter 2 - Review of the Literature

### General Theoretical Considerations

A major problem confronting those who seek to study the phenomenon of creativity is that of definition. The question of appropriate criteria for labeling something creative is as yet unresolved, as the following comment illustrates:

> A striking feature of the literature on creativity is the diversity of interests, motives, and approaches characteristic of the many investigators. Creativity has been viewed as a normally distributed trait, an aptitude of life. It has been described as that which is seen in all children, but few adults. It has been described in fine arts, or new thoughts. Creativity has been described as being caused by self-actualization and by sublimation and restitution of destructive impulses. Clearly there is a need for organization and integration within the psychological study of creativity. What are the many investigators studying? How are they studying it? Four contemporary emphases are apparent: products, process, measurement, and personality. (Golann, 1963, p. 548)

These emphases or combinations of them are a basic determinant of any individual definition of creativity. Another such determinant

would be the basic theoretical source of the conception of creativity. Common examples of such theoretical sources would include: traditional logic, associationism, Gestalt theory, psychoanalytical and Neo-Freudian conceptions, "dynamic perception" theory, non-Freudian personality theory, and psychological traits theory, to name several. These different theoretical sources and the various psychological approaches to the construct of creativity indicate the complexity of the task of finding a commonly accepted definition of creativity.

The task of empirically studying creative behavior requires some form of measurement which precludes an operational definition of creativity as a function of the assessment techniques utilized. This operational definition must serve as a transition from the theoretical consideration or definition to the applied rationale of the measurement devices. This operational definition and related rationale will be discussed in relation to the description of the measurement tasks in Chapter Three. The theoretical conception employed, in addition to generating the operational definition, also becomes the focal point for considering the relevant literature.

To establish this focal point this study utilizes definitions selected from the hundreds that exist in the literature. The first of these has been offered by I. A. Taylor (1959) who proposed five levels of creativity. This definition was developed by analyzing over 100 different definitions of creativity. The following represents a summary of this definition made by Golann (1963):

> Expressive creativity is most fundamental, according to Taylor, involving independent expression where skills, originality, and the quality of the product are unimportant. Spontaneity and freedom are apparent from which later creative talents develop. Individuals proceed from the expressive to the productive level of creativity when skills are developed to produce finished works. The product is creative in that a new level of accomplishment is reached by the person though the product may not be stylistically discernible

> from the work of others. Inventive creativity is operative when ingenuity is displayed. This level involves flexibility in perceiving new and unusual relationships between previously separate parts. It does not contribute to new ideas but to new uses of old parts. Innovative creativity requires strong abstract conceptualizing skill and is seen when basic foundation principles are sufficiently understood so as to allow improvement through modification. The highest form of creativity is emergentive creativity, which involves the conception of an entirely new principle at a most fundamental and abstract level. (pp. 550-551)

This definition seems to synthesize the breach between the product-oriented investigation of creativity and the orientation focused on analysis of processes involved in creative thinking or production. This synthesis is on the basis of developmental levels of creativity as reflected in Taylor's five levels. This synthesis is significant because if the nature of the development of creative capacity is to be investigated, some means of assessing development at the first two levels is desirable, and yet unattainable through a traditionally product-oriented approach. The problem becomes one of assessing behavior in such a way as to be able to make inferences as to the extent or degree of capacity present in any given individual.

Such assessment efforts have largely been made possible through the work of J. P. Guilford and E. Paul Torrance. The work of both of these men has been based on the assumption that creativity can most productively be viewed as the outcome of a combination of psychological traits. Guilford has proceeded in an attempt to demonstrate the presence of such traits through factor analysis and both Guilford and Torrance have developed instruments to measure performance on tasks related to these traits.

The following represents a summary of the traits that Guilford believes to be related to creativity and the measuring devices he

has utilized to assess these traits. The similarity of these traits and tasks to the first three levels of creativity defined by Taylor is significant.

The traits Guilford relates to creativity (Guilford, 1956, 1957, 1959a, 1959b, 1959c; Guilford, Kettner, and Christensen, 1954, 1956) are described as: ability to see problems, fluency of thinking (both the factors of word fluency and ideational fluency), flexibility of thinking (both the factors of spontaneous flexibility and adaptive flexibility), originality, redefinition, and elaboration. Golann (1963) has summarized from the above works of Guilford, et al., the following tasks as characteristic of the measuring devices designed or adapted to assess these traits:

> Very briefly described, his tests require individuals to state defects or deficiencies in common implements or institutions; to produce words containing a specified letter or combination of letters; to produce in a limited time as many synonyms as they can for a stimulus word; to produce phrases or sentences; to name objects with certain properties (for example objects that are hard, white, and edible); or to give various uses for a common object. (p. 553)

Torrance (1962) the author of the creative tasks this study will use, has defined creative thinking:

> ...as the process of sensing gaps or disturbing, missing elements; forming ideas or hypotheses concerning them; testing these hypotheses; and communicating the results, possibly modifying and retesting the hypotheses. (p. 16)

Torrance continues:

> In accepting this kind of definition of creativity, a variety of kinds of behavior are included. It is my opinion that we must continue to do this. By this, it is not meant that we should try to represent

> all of these abilities and/or behaviors by any single index. Neither does it mean that we are now ready to establish a set of discreet abilities or pure factors. (p. 17)

These definitions and concepts then are the base of the theoretical construct of creativity and its component traits which will be investigated in this study.

There is a large body of theoretical and empirical literature which analyzes the determinants of creative behavior, as defined by Taylor, Torrance and Guilford. The rationale for the hypotheses suggested in Chapter One derives from that literature. For convenience these determinants can be divided into two major categories, although there is considerable overlap between them. The first of these is school experience as represented by both the objectives and the methods of formal schooling. These are perceived generally as detrimental to creative development and expression. The second category represents a more general societal force, the very process of socialization. A variety of evidence can be cited to show that a significant proportion of what is done in the name of "preparing individuals to fit into their role in society" is detrimental to creative development and expression. The formal process of education is one of the major socializing institutions in our society and it would seem logical to conclude that the pressures exerted by this institution would be significantly less for those whose involvement in formal school experiences had been limited.

## Effects of Socialization Process

The first considerations are the general, or more pervasive, factors of the socialization process. Harold Anderson (1961) has described and distinguished between the "Open System" and the

"Closed System", within which an individual interacts with his environment. The "Open System", which is conducive to creative expression and development, is characteristic of infancy and early childhood. During this time a child confronts more and more frequently a system of environmental demands, taboos, socializing and enculturating processes which transform his relation to his environment into a "Closed System," which is not conducive to creative expression. Anderson depicts exposure to a school curriculum, which is preponderantly a "Closed System," as a major force of the transformation process. Torrance (1962a) has pointed out that society in general, is "downright savage" towards creative thinkers, especially when they are young. He points out that the high degree to which school administrators and teachers are coercive and emphatic in the establishment of behavioral norms is a function of societal expectations and pressures.

Across all theoretical descriptions of the creative process there is marked emphasis upon avoidance of rubricizing and stress upon perceiving phenomena in their unique concrete totality rather than in stereotypes or cliches, of seeing beyond the familiar or conventional interrelationships of experience and phenomena and perceiving new relationships. Schachtel (1959) has noted that cultures vary in the degree to which they impose stereotypes on experiences and memory, and states that Western civilization in the last one hundred years, has been characterized by imposing conventional schematization on experience and memory at an accelerated rate.

The need for social relationship has manifested itself in American Society to such a degree that anthropologists refer to our society as the most peer-oriented culture in the world. The references to the inhibiting effects of peer pressure on creative expression are numerous in the literature. Torrance (1962a, 1963b)

has empirically demonstrated the inhibitory effect on pre-identified creative thinkers by peer pressures in group problem-solving experiments. In addition he has shown the intensity of the group's rejection of both the ideas of the creative individual and of the individual himself. Yamamoto (1960) has demonstrated the same tendency and has shown how sociometric choice of creative individuals decreases with age as more and more weight is placed on the "social" evaluation of behavior.

Torrance (1963a) reports that students and investigators of creative development have observed a crucial fourth grade slump and a less severe, but still significant slump, at the time of grade seven. Torrance has documented this observation with creative thinking tests. Although there are numerous causal factors involved in each of these slumps, a major cause is the heightened concern at these ages for conformity to different types of peer pressures at different ages. Therefore many young people give up their creative pursuits for conventional involvements. By finding no similar discontinuities in the task performance of children of other cultures and societies Torrance has identified these developmental discontinuities in American culture as man-made: the product of cultural forces.

Another reason for the "fourth grade slump" has been posited by Torrance (1963b, 1963c). This factor is one of proper sex-role behavior. McKinnon (1962), Barron (1955), among others, have demonstrated in various ways that creative people do not conform to stereotyped cultural sex-roles, and these authors describe such conforming as a hindrance to creative expression.

Another strong cultural or societal attitude which is suppressive of creative expression or development is the attitude manifest towards divergency. Once even leading thinkers themselves believed "genius" and "madness" to be closely associated. Although

these notions have been discredited, somehow the belief has persisted that any divergence from normalacy is unhealthy, and is to be corrected at all costs. Despite the increasing prevalance of theories of psychotherapy which stress self-fulfillment and creative living, divergency from the norm is still regarded with suspicion. Parents prefer having their children "well adjusted," rather than "different." Torrance (1962a) analyzing imaginative stories composed by elementary school children, concludes that they are taught very early that divergent characteristics or behavior are signs of mental illness.

Another normal aspect of childhood thinking, fantasy, is also a target for early elimination in our society. Although the relationship between fantasy and creativity varies significantly according to theoretical positions, Torrance (1963b) has cautioned that we need to keep fantasy alive until the child's mental development is such that he can engage in realistic creative thinking. The parental and societal pressures against fantasy are currently not timed harmoniously with this goal. Testing of first and second graders has indicated that many children with impoverished imaginations have been subjected to some rather vigorous and stern efforts to eliminate fantasy too early. (Torrance 1963b)

Earlier reference was made to Schachtel's (1959) comments on the high degree of conventionalizing of experience and memory schemata in Western Civilization. At this point it is necessary to consider the consequences of this in more detail. Schachtel's notions give us a general description of the effects of socialization. The various factors discussed above, each in their own way, contribute to this general effect. Schachtel points out that:

> No greater change in the needs of man occurs than that which takes place between early childhood and adulthood. Into this change have gone all the decisive formative influences of the culture transmitted

> by the parents, laying the fundament of the transformation into the grown-up, 'useful' member of society from the little heathen, who is helpless but as yet sees nothing wrong with following the pleasure principle completely and immediately and who has an insatiable curiosity and capacity for experience. (p. 284)

It should be noted that just such an insatiable curiosity and capacity for experience are commonly held to be necessary for any degree of creative expression. But, as Schachtel points out:

> In the course of later childhood, adolescence, and adult life, perception and experience themselves develop increasingly into the rubber stamps of conventional cliches. The capacity to see and feel what is there gives way to the tendency to see and feel what one expects to see and feel because everybody else does. Experience increasingly assumes the form of the cliche under which it will be recalled because this cliche is what conventionally is remembered by others. This is not the remembered situation itself, but the words which are customarily used to indicate this situation and the reactions which it is supposed to evoke. (p. 288)

Rugg (1963) has also called attention to this attribute of development, using the concept of stereotyped perception.

Both Rugg and Schachtel have emphasized the role of language development as an implement for the enculturation of memory and experience. This is of particular significance for this paper since a major emphasis during early years of the formal educative process is language development.

Schachtel (1959) points out that language is the major form or determinant of the memory schemata and has an obscuring function. Even the adult encounters the problem of the incompatibility of

of experience with language and the consequent forgetting or distortion of experience by the cliches of language.

> It is chiefly during the period of early childhood that the quality of the world around him changes for the growing child from a place where everything is new and to be explored — to be tasted, smelled, touched and handled, wondered about and marveled at — to a place where everything either has received a name and a label or is potentially capable of being 'explained' by such a label, a process which will be systematically pursued in school. No experience, no object, perceived with the quality of freshness, of something wonderful, can be preserved and recalled by the conventional concept of that object as designated in its conventional name in language. (p. 293)

It would seem obvious that Schachtel, along with many others, finds education guilty of failing to sustain or develop the desirable qualities of childhood. He refers to the curious, spontaneous, pleasure seeking child clashing with the repressive, conventionalized, pleasure forbidding culture in a "battle" called education, a battle in which the child is always the loser.

## Effects of Formal School Experience

Why must the child or his creative expression be the loser in this clash called education? What are the conditions specific to the formal educative process that have resulted in current literature leveling such a widespread accusation that education is a suppressant of creative expression and development?

One way of attempting to answer this question is simply to evaluate the schools' role with respect to the conditions just discussed, conditions attributed to the pervasive influence of culture

and society. These conditions would include such factors as premature attempts to eliminate fantasy, restrictions of curiosity, overemphasis or misplaced emphasis on sex roles, overemphasis on prevention of failure and mistakes, and misplaced or mistimed emphasis of certain types of verbal skills. At best American education can be characterized as doing little to actively foster or sustain creative thinking. At worst the schools can be accused of actively fostering and strengthening the conditions out of which these obstacles emerge.

Earlier the authoritarian nature of our culture and society was alluded to as a suppressant of creativity. This is particularly manifest in the objectives and methods of formal education. Anderson (1961), Torrance (1963a), and Schachtel (1959) have all referred to the creative process as the natural way of learning. Torrance points out that man prefers to learn by exploring, manipulating, questioning, experimenting, risking, testing, and modifying ideas. Yet we have habitually insisted that people learn by authority, using the rationale that learning this way is more efficient and more economical. Torrance (1963b) places part of the blame on teacher training practices.

> Colleges and universities are still teaching only the psychology of learning, almost never the psychology of thinking. Teachers of the near future are still being taught only how to construct tests to assess what pupils learn, almost never to determine what pupils can do with what they learn. (p. 4)

Anderson (1961) has depicted the "closed system" of education as concerned with memorizing facts, formulas, and beliefs, and the acquiring and storing of information. Such information may be well verified scientific findings, or may easily be biases, prejudices, or superstitions. Schachtel (1959) has described the

developmental conflict of natural learning processes and education's role in the following way:

> It is safe to assume that early childhood is the period of human life which is richest in experience. Everything is new to the newborn child. His gradual grasp of his environment and of the world around him are discoveries which, in experimental scope and quality, go far beyond any discovery that the most adventurous and daring explorer will ever make in his adult life. No Columbus, no Marco Polo has ever seen stranger and more fascinating and thoroughly absorbing sights than the child that learns to perceive, to taste, to smell, to touch, to hear and see, and to use his body, his senses, and his mind. No wonder that the child shows an insatiable curiosity. He has the whole world to discover. Education and learning, while on the one hand furthering this process of discovery, on the other hand gradually break and finally stop it completely. There are relatively few adults who are fortunate enough to have retained something of the child's curiosity, his capacity for questioning and for wondering. The average adult 'knows all the answers.' Which is exactly why he will never know even a single answer. He has ceased to wonder, to discover. He knows his way around, and it is indeed a way around and around the same conventional pattern, in which everything is familiar and nothing cause for wonder. It is this adult who answers the child's questions and, in answering, fails to answer them but instead acquaints the child with the conventional patterns of his civilization, which effectively close up the asking mouth and shut the wondering eye. Franz Kafka once formulated this aspect of education by saying that 'probably all education is but two things, first, parrying of the ignorant children's impetuous assault on the truth, and, second, gentle, imperceptible, step-by-step initiation of the humiliated children into the lie'. (pp. 292-293)

Authoritarianism is also the dominant characteristic of the relationship that typically exists between student and teacher. The consequences of the nature of this relationship have been investigated by Torrance (1962a) and Getzels and Jackson (1962). Both

investigators reported that teachers rated the highly creative children, in comparison with the highly intelligent ones, as less desirable as pupils, as not as well known or understood by them, or less ambitious, and less hard working or studious. Both found significant differences between values of the creative group and the values of the teacher. Both studies showed no significant difference in achievement, as measured by standardized achievement test scores between the high creative and the high intelligence group; yet the latter group, the highly intelligent, consistently received higher teacher evaluation. Elizabeth Drews (1961) found that of three types of gifted high school students, the studious, the social leaders, and the creative intellectuals, the poorest teacher grades were made by the creative students. However, in competitive examination sampling a wide range of information, they performed better than either of the other two groups.

Torrance (1963b) suggests that a major reason for the reluctance of elementary and secondary teachers to give attention to the development of creative capacities is the admissions policy of colleges and universities, especially the highly "convergent-oriented" admissions tests. Getzels (1960) has pointed out that the tests, recommendations, and rank in class relied upon so heavily for college admission are biased in favor of the student with "convergent" intellectual ability and social interests. The fact that the vast majority of scholarship decisions are made on identical criteria, adds strength to this creativity deterrent.

A maxim in American education for a long time has been that most administrators evaluate teachers on the criteria of the control they exercise over their class. Thus the maintenance of expected levels of classroom control are vital to the teacher. Torrance (1962a) has discussed various underlying, basic strategies

employed by teachers to maintain this control and has noted that these strategies are generally detrimental to creative development. Jules Henry (1959) has described a growing trend in classroom control techniques:

> The contemporary idea that good elementary school teachers should be accepting, giving parents has resulted, as we have seen, in the teacher's using affection as a defense against other impulses of the children: the teacher stimulates their love by calling them 'honey' and 'dear' and by fondling, while at the same time she awakens fear of loss of her love if they get out of line. (p. 269)

The dynamics of control lie in the use of guilt, in making the child aware of the fact that he has disappointed the teacher. Henry's description is almost synonymous of that given by Urie Bronfenbrenner (1961) describing the dominant method of control used by parents. He refers to this as the "love-oriented" method of child rearing. Its dynamics are, similar to what Henry describes, the withdrawal of love by warm, loving parents, and the capitalization on guilt. Although Bronfenbrenner finds children raised by this technique to possess such culturally valued characteristics as self-control, achievement, responsibility, leadership, popularity, and adjustment, he also reports these children as being more anxious and dependent, and lower in such qualities as independence, initiative, and self-sufficiency. These characteristics, present or lacking, concur significantly with both Getzels and Jackson's (1962) and Torrance's (1962a) descriptions of high intelligence-low creative and high creative but of comparably lower, but still high intelligence, children.

Frequent mention has been made of the consequences of the generally negative or low evaluation made of creative individuals by teachers, in an informal or global sense. It is also necessary

to look at the consequences of evaluation techniques, not of individuals but of their performance, that are dominant in the educative process. One aspect of this area of concern is the instruments of measurement utilized by educators as the basic criteria in making evaluative judgements. Such instruments run the gamut of crude, quickly thrown together teacher-made tests, through those tests published to accompany text books, to the standardized tests prepared and sold by test publishers. These tests predominantly show one quality, the reliance upon recognition responses. This excessive reliance limits sampling of the more complex outcomes of schooling. Even the best batteries of standardized tests show few items exceeding level 4 in Bloom's Taxonomy of Educational Objectives. (Torrance, 1962a) Similarly in the area of ability testing, the realm of divergent thinking abilities goes untapped because of unsuitability to the recognitive format. Anderson (1961) has stated that the goal of the testing movement was essentially correlation and prediction.

> But only similarities can be correlated, and only the defined can be predicted. The uniqueness, which can neither be predicted nor in advance defined, and which is a necessary evidence of creativity has been generally ignored or discarded by the statisticians and the testers. Only recently have there been efforts to capture and assess in test responses qualities of originality which for forty years have been discarded. (p. 12)

Unfortunately, it might be added, that such recent efforts as Anderson notes have been primarily concentrated on specific tests of "creativity" and have not, to any significant degree, been integrated into measurement of achievement or ability.

The most frequently used form of test item is the multiple choice item. Although experts of test construction have demonstrated that this form can assess any level of learning or

thinking, the vast majority of items used for measurement and subsequent evaluation fall into the recall or recognitive range. Such is also true of true-false items, fill in the blank, and most short-answer items. A majority of essay items do not in any real sense transcend these levels of comprehension. Both professional and popular literature in the last ten years has found frequent reference to the failings of such measurement techniques, and specifically their failure to assess or reward the creative thinker. (Taylor, 1960; McNeil, 1960; Strang, 1959; Vernon, 1960; Guilford, 1959; LaBrant, 1959; Morgan, 1960; Whyte, 1956; Barzun, 1959; and Hoffman, 1961)

It must also be considered that standardized achievement tests have other influences that, because of the limitations of the tests, can be non-supportive of creative development. Such tests strongly influence the emerging objectives of education through several factors. First, there is heavy public pressure for the schools to realize high scores on these tests. Also the widespread use of such tests for decisions relative to grouping, college admissions, and identification of the talented for special programs, give these tests and their implicit objectives a pervasive influence on education.

Such decisions are even more heavily rooted in the individual's performance on the standardized intelligence tests. Getzels and Jackson (1962a) identified the degree of significance that has been granted the I.Q. concept:

> The concept of intelligence and the consequent intelligence measure have been used to define individual difference in cognition as if the concept and the measure encompassed the totality of the human mind and imagination. In schools -- and more recently in other areas requiring intellectual accomplishment -- the I.Q. (or some cognate of it) has become the critical metric on which individuals are

> evaluated and sorted, given preference or denied it. Individual differences in potential for productive thinking have been made synonymous with individual differences on one or another of the numerous intelligence tests. (p. vii)

This statement reflects the authors' criticism of the uses of and inferences made from intelligence tests, not the tests themselves. Others tend to be more critical of the tests and their underlying constructs. (I. A. Taylor, 1959; C. W. Taylor, 1959; Smillie, 1959; Torrance, 1962, 1963b; Getzels and Jackson, 1962; McKinnon, 1962; Barron, 1955; and others)

These are comments and observations concerned with the formal process of evaluation in our schools. At the same time there is concern over the informal process of evaluation of school work and the products of that work. The formal instruments of measurement are used at certain points to assess achievement, but pervasively, day by day, hour by hour, teachers and for that matter parents, in their supervisory role make evaluations of the efforts and achievements of children. Torrance (1963a) has stated that we need to provide chances for children to learn and discover without immediate evaluation. Constant evaluation, he claims, tends to make children afraid to learn creatively. However, constant evaluation, or over-evaluation, is characteristic of adult-child relationships in our society. The primary consequence of this evaluation is that far too many children fail to develop realistic self-concepts, fail to test and, therefore, fail to define their limits. Children are not provided with situations in which it is safe to practice without evaluation. They are aware that their performance or production will seldom measure up to those of adults and, therefore, fear rejection and failure. It is safer to stay within the known areas of adequacy and maximize the probability of success. The degree of risk-taking that individuals are willing to demonstrate in

their exploration and thinking diminishes with age (Torrance 1962a). (Schenitski (1961),using two of Torrance's tasks of creativity, assessed the effect of evaluation versus no evaluation during practice on subsequent test performance. In general, the unevaluated practice was more effective than the evaluated practice in encouraging creativity in the first three grades but not in the upper grades.

Wallach and Kogan (1964):

> ...have emphasized that those psychological processes associated with creative functioning require, for their optimal operation, a context free from or minimally influenced by the stresses that arise from academic evaluation and a fear of the consequences of error. To further this kind of goal within education, then, is to fashion a learning and teaching environment that will permit children to minimize the bind produced by negative sanctions for error. To be sure, it would be desirable to further such a goal with regard to the traditional domain of intelligence, too. (p. 320)

As has been implied in the comments on evaluation, a source of much of the problem is the objectives set by schools and by the individual teachers. Torrance (1963b) has dealt directly with the matter in a study related to the teaching of social studies, a curriculum area generally regarded as relatively free from the more closed, stereotyped traditional values and objectives:

> A questionnaire sent to a random sample of elementary and secondary social science teachers asked them to name a subject or unit they teach in social studies and then list what they considered the three most important objectives of this subject or unit. Each objective was then classified according to the type of mental operation pupils would have to engage in to achieve it.

> The results show that an overwhelming majority of the social studies objectives fall into the cognitive category, over 70 per cent for the entire sample. Most of the remaining objectives require conformity to the behavioral norms -- about 17 per cent at the elementary level and about 21 per cent at the secondary level. Eliminating the five per cent that require memory, precious few remain for either of the thinking categories. (pp. 4-5)

Although this research covered only one area of curriculum, the nature of the stated objectives seems typical for school curricula generally. Frequently, if not always, the school's philosophy and stated objectives are replete with reference to thinking, judging, creating, etc., but when it comes to what is rewarded by the school, the emphasis documented above by Torrance is a far more accurate picture.

There are many pressures in our schools and society in general which push the creative individual towards the mean, towards the average. A number of these pressures work not by direct coercive force, but rather by establishing situations which fail to reward the individual for uniqueness, his diversity, his creativeness. Wolfe (1960) cites as examples the use of uniform assignments, methods of selecting students for the next higher educational level or for scholarships, advertising procedures, trade union policies, wage scales, and such. Torrance (1952a) and Getzels and Jackson (1962) all showed that the special programs for the talented, certainly a reward to participating individuals, would exclude seventy percent of the most creative individuals when based on the typical cut off points on intelligence tests. American education has thus far paid some lip-service to creativity as an objective, but has done little to change evaluation techniques to reward creative behavior.

Grover (1963), Barron (1955), and others have implied that, in part, creativity is seldom rewarded, or even recognized, because of an emphasis on product rather than process. This is well illustrated in education's approach to problem-solving tasks. The development of creative thinking with respect to problem-solving is in a large part dependent upon developing a sensitivity to problems. This is emphasized in Torrance's definition of creativity and in the tasks designed by Guilford to measure creative traits, both quoted earlier. Rugg (1963) believes that an incomplete interpretation of John Dewey has, for fifty years, set education on the track of emphasizing problem-solving and neglecting the prior phase -- the discovery of the problem. Anderson (1961) has criticized the approach to problem-solving for a different reason. He feels that, "In the Closed System of education problems have 'fixed answers', that is, the answers are in the back of the book, they have been agreed upon by the culture, or they are found in the teacher's head or in the program of the teaching machine." Getzels (1964), while also dealing with sensing problems, and of predetermined answers, adds a third dimension, which is a known or fixed method for solving the problem. He points out that problem-solving tasks vary by combinations of these variables as either known or predetermined or unknown, with no fixed answer or method of solving. He observes that a substantial majority of activities labeled as problem-solving by our schools deal with pre-formulated problems, with one known and determined correct method of solution and with a known and fixed answer. This he concludes is not adequate for developing creative problem solvers.

## Summary

The general summary or synthesis of these considerations which serves as the rationale for the predictions to be made in the study

must emphasize two factors. The first of these is the negative or suppressing role that education plays with respect to many of the psychological characteristics essential for creative expression. At best education is seen by some as taking a relatively passive role, perhaps not always suppressing, but seldom sustaining or rewarding creative endeavors. The above rationale has attempted to establish this role as being a consequence of the various methods used in the schools and of the objectives, both explicit and implicit, of American formal education. It was further emphasized that a major objective of education is socialization and that this is not conducive to sustaining creative capacities in individuals. Secondly, the above evidence would seem to indicate that certain years, or groups of years are more crucial than others. More specifically, there would seem to be grounds for assuming that the early years of formal schooling are the most crucial. This would be supported both by the numerous references quoted above and found throughout the relevant literature on the creative nature of the experiences and characteristics of early childhood. Where early childhood ends and another stage begins is not clear; but the strong empirical evidence of the so-called fourth grade slump would give some indication that, by this time, early childhood has passed into a more "advanced" stage of development. If this assumption is correct, a variation in involvement in formal education during the first three years should generate greater performance differences on tasks of creativity than will variations after the third grade. The data collected from this study should allow for a tentative examination of this assumption as well as the more general hypothesis that students whose exposure to formal schooling has been less than that of their age-peers, will achieve higher performance scores on tests of creative thinking.

## Chapter 3 - Procedures

### Population and Sample

Samples from two populations were utilized for this study: (1) Negro students attending the schools of the Free School Association in Prince Edward County, Virginia and (2) Negro students in the public school system in Jackson, Michigan. Population (2) is comprised of children whose involvement in formal education has been of the typical length and regularity for American children. Population (1) is comprised of students whose involvement in formal education has been delayed or interrupted for a period of years. The length of the delay or interruption is dependent upon their age.

The samples from the two populations were composed of children of four age levels representing first, fourth, sixth, and eighth grade students. The rationale for this age selection represents an attempt to study the differential effect of the lack of formal schooling in a quasi-developmental context. Although it would have been more desirable to test all ages consecutively, time limitations made this impossible and a sample of every other year was utilized,

with the exception of the three year gap between first grade and fourth grade.

First grade students will serve as a control factor for cultural variation. The first grade students in both populations will have had essentially one year of formal schooling at the time of the testing. This is assuming that kindergarten experience, unique to the Jackson, Michigan children is not a significant variable. If the hypothesis is verified that there is no significant difference in task performance between the samples at this age level, this will serve to validate the inference that differences found at the other age levels will be substantially a function of variation in involvement in the formal educative process.

Because the time factor limited the investigation to an adequate sampling of four age levels, the other three age groups were chosen with respect to variations in school involvement during the primary years. Thus the fourth grade students in population (1) are completing their first year of formal schooling, the students in population (2) their fourth. Torrance (1962a, 1963b) has pointed out and documented that fourth grade represents the period when the most significant slump or repression of creative behavior occurs in the developmental pattern of American children. As discussed earlier, the basic causes for this slump are indigenous to the socialization process, a process implemented to a significant degree by formal education. Thus students with minimal exposure to formal education should have retained a higher potential and functional degree of creative capacity.

Both the sixth grade students and the eighth grade students in population (1) missed four years of school. However, the former had one year of school, then a four year interruption, and then one

more year; whereas the eighth grade students had three years of schooling, missed four, and then had one more year. The students in population (2) had six and eight years of schooling, respectively, as compared to two and four years for the students in population (1). If the early elementary years are the most crucial with respect to the role of education in the development of creativity, the performance scores of the eighth graders in the two populations would show a higher similarity than at grades four and six.

The samples from each population were designed to consist of 24 students, 12 boys and 12 girls, at each of the four age levels. As the data will disclose this was not achieved for all groups. The students were selected from among those in the prescribed grades on no other basis than having birthdates within a six month age range. Thus for the first grade the range is from 3/1/57 to 9/1/57; for the fourth grade the range is 3/1/54 to 9/1/54; for the sixth grade, 3/1/52 to 9/1/52; and for the eighth grade 3/1/50 to 9/1/50.

It was further necessary to specify for the Prince Edward County sample, that these be students who did not attend any school during the four year period when the public schools in the county were closed. This was necessary as a number of the students had attended schools elsewhere in the nation for one to four years.

The sampling of these two populations to test the hypotheses suggested earlier and specified later in this chapter raises a significant question. Can it be assumed that the two populations, Prince Edward County, Virginia and Jackson, Michigan are basically similar on other variables that might be deterministic of performance on creative tasks other than that variable or set of variables labeled as formal school experience? A major difference is quite apparent, the contrast between northern-urban and southern-rural.

Numerous other differences which may or may not exist would include such a matrix of factors grouped as family patterns, religious affiliations, socio-economic strata and its differential implications in the two settings, general community attitudes, and others. These factors were not controlled for and must be considered serious limitations of the study. Only race and school experience were controlled. The original plans for the study had included a similar sample drawn from a neighboring county in Virginia. Had it been possible to collect the same array of data on this group, this particular limitation of the study would have been minimized.

## Data and Instrumentation

As should be noted in the title, this study does not claim to be studying the effect of lack of formal schooling on creativity, per se, but rather the effect on performance on creative tasks. It is not claimed by way of an operational definition that the instruments utilized measure creativity, but rather, consistent with the psychological trait approach which gave rise to these tests, that the tasks measure certain intellectual traits which are necessary, but not sufficient for, creative thinking or expression. Such a definition is consistent with the five-level definition of Taylor quoted earlier on page five (5). It is consistent in the sense that it measures expressions of creative ability dependent upon the characteristics and capacities described in the first two and possibly three levels given by Taylor. It thus focuses on inferring from response behavior the comparative degree to which the individual is able to demonstrate the trait being assessed, rather than relying on a more subjective judgement as to whether or not a product is creative. The latter form of evaluation, besides being heavily influenced by personal and/or social values, is more appropriate for

the products suggested by levels four and five in the theoretical definition. Developmentally, this is also a significant consideration because of the ages of the children being tested. It would not be expected that children of these ages would have developed intellectually to the point of performing at levels four and five. High performance on these tests would indicate a trait capacity which would make it possible for them to eventually achieve at these higher levels, but does not necessarily guarantee or even predict such to be the case. However, low performance by an individual would cause one to predict that the higher levels of creative achievement would not be reached by such individuals.

Performance on measures of four traits identified as crucial to creative thinking by Guilford (1956, 1957, 1959a, 1959c) and Torrance (1962a), was assessed. These four traits are: (1) fluency; (2) flexibility; (3) originality; and (4) elaboration. The definition of these traits can best be developed in conjunction with the scoring procedures for the different tasks used.

Four different tasks were utilized to assess these traits. These tasks were chosen from a wide variety of tasks used for studies of creativity by E. Paul Torrance of the University of Minnesota (1962a). The tasks used are identified as Abbreviated Form VII, Minnesota Tests of Creative Thinking. The form is comprised of four tasks, Incomplete Figures, Circles, Product Improvement, and Unusual Uses of Tin Cans. Torrance (1963d) stated:

> These four tasks have been selected from a variety of tasks which have been developed during the past five years by the author (Torrance). They have been selected as the four tasks which in combination might yield

> the greatest amount of information concerning creative growth. (p. 2)

In addition to the fact that these tests measure traits consistent with the theoretical definitions of creativity posited for this study, the basis for the selection of this particular battery of tasks was influenced by three other factors. First, it was necessary to employ tasks which were applicable to a relatively wide range of ages. These tasks meet this criterion and furthermore, Torrance's tasks represent the only fairly well standardized battery available for the lower or primary grades. Second, it was necessary to select tasks utilizing stimuli which would minimize any effect on performance generated by socio-cultural differences. Third, the choice was influenced by a desire to vary the tasks with respect to their dependency upon verbal skills. The selected battery contains two primarily non-verbal tasks and two verbal tasks. Torrance (1962d) has stated:

> This enables us to study both verbal and non-verbal kinds of behavior. Some individuals, we find, are quite bold and imaginative when they are allowed to express their ideas non-verbally but are tremendously constricted, almost paralyzed, if they have to express their ideas verbally. For others the reverse is true. (p. 4)

Although the inclusion of both verbal and non-verbal tasks allows for a wider variety of individual differences of capacities to be tested, the major consideration revolved around the effect of lack of formal schooling. Such a lack, in its particular way, represents a form of deprivation, and the literature on academic deprivation, as a specific form of cultural deprivation, has established the significantly lower verbal skills of these children in comparison with the normal or "non-disadvantaged" youth. Thus, it was necessary to provide a non-verbal

form of creative expression.

The following represents a description of the tasks (Torrance 1962a, 1963d):

Nonverbal Tasks

Incomplete Figures

This test is an adaptation of the Drawing Completion Test developed by Kate Franck and used in the studies of creativity by Barron (1958) and others. The first page consists of instructions and four squares, each containing a stimulus. The dimensions of the squares are nine square inches. The second page contains six such stimuli squares. The instructions are as follows:

> By adding lines to the figures on this and the next page, you can sketch some interesting objects or pictures. Try to think of some picture or object that no one else will think of. Try to make it tell as complete and as interesting a story as you can by adding to and building up your first idea. Make up a title for each of your drawings and write at the bottom of each block next to the number of the figure.

It should be noted that these directions as well as those for the other three tasks, were found to be adequate for the students, with some repetition and slight variation for the various ages. This was established as a result of a pilot study conducted by the investigator in Prince Edward County, during which the battery was administered to a total of sixteen children representing grade levels first, fourth, and sixth. Another point which merits attention is the requirement in this task and the other nonverbal one, Circles, of naming the drawing made from the stimulus. As might be suspected

Guilford, (Guilford, Ketter, and Christiansen, 1964) has demonstrated that this represents a naming or labeling factor which is strongly related to verbal comprehension. Although this makes the task less than completely non-verbal, the tasks are less dependent upon capacity in verbal skills than are the two tasks identified as verbal tasks.

The rationale of this test, as explained by Torrance, is based upon incomplete figures creating tensions in the individual with the usual response being to complete it in the simplest and easiest way possible. Thus the subject has to be able to handle his tensions and delay the gratification of this impulse in order to produce an original and elaborate set of figures.

The scoring for this test, as for the other three, has been well defined in the test manual. Fluency for all four tasks is defined as the total number of relevant responses, relevancy being defined in terms of the task assigned. In other words, the fluency score for Incomplete Figures is the number of completed figures. Flexibility is obtained by counting the number of different categories of objects the completed figures represent. There are seventy such categories listed in the manual. Examples of such categories are:

1. Aircraft: airplane, bombers, rockets, space ships, etc.

2. Animal: including animal faces, heads, and tracks; ape, bear, bull, cat, camel, deer, dog, elephant, goat, hog, horse, mouse, snail. Also include specific breeds, as dogs (collie, poodle, etc.) or cats (Siamese, tabby, etc.)

3. Building: animal house, apartment house, bee house, church, fort, home, house, oriental house, pagoda, store, temple, etc.

4. Human Form: human figure, including face, person, a specific person such as Mitch Miller, woman, dancer, clown, cowboy, etc.

5. Time Piece: clock, hour glass, sand clock, watch, etc.

The flexibility score for the test is the sum total of different categories used in the responses. For example, if an individual completed the first stimulus in such a way as to draw a house, any form of building drawn on responses 2-9 would not add to the flexibility score.

Originality for this task, as in the other three, is defined as "uncommonness of response" in a statistical sense. A separate scoring guide is provided for each of the ten figures presented, since each tends to elicit different responses. The frequency norms are based on a tabulation of the responses of 1,056 subjects from first grade through high school. Responses having a frequency of 5 per cent or more are given 0 credits, and are listed in the guide. Responses with a norm frequency of 2 per cent to 4.99 per cent are given 1 point and are also listed in the guide. The scoring instructions then stipulate that all other responses showing imagination and creative strength are to be given 2 points. It thus becomes necessary at times to make judgements about unlisted responses; although in fact, these are relatively scarce. The procedure followed was to limit non-imaginative, unlisted, responses that are characteristiced by simple closure of the stimulus to one point; all others received two points.

The scoring of elaboration is based on the rationale that the minimum and primary response to the stimulus figure is a single response and that the integration and exposition of detail beyond this single response is a function of creative ability. Therefore, in scoring elaboration, credit (one point) is given for every

pertinent detail (idea) added to the original stimulus figure itself, to its boundaries and/or to the surrounding space. However, the basic response itself must be meaningful before elaboration can be scored. One point is given for: (1) Each essential detail of the responses, but once that class of detail is scored, further evidence of the same class is not counted. (In other words, each additional idea that contributes to the story the picture tells is credited with a point, but the repetition of an idea does not count.) (2) Deliberate shading (not just going over the same lines again). (3) Decoration, only when meant as such. (4) Each major variation (not of quantity) of a design which is meaningful with reference to the total response. (5) Each elaboration of the title beyond the minimum descriptive label.

## Circles

The Circles Task was developed by Torrance and his associates. The task consists of a first page with a set of instructions plus six circles the size of quarters. The second page contains thirty circles of the same size. The instructions are:

> In ten minutes see how many objects or pictures you can make from the circles below and on the next page. The circles should be the main part of whatever you make. With pencil or crayon add lines to the circles to complete you picture. You can place marks inside the circles, outside the circles, or both inside and outside the circles - wherever you want to in order to make your picture. Try to think of things that no one else will think of. Make as many different pictures as you can and put in as many ideas as you can in each. Make them tell as complete and as interesting a story as you can. Add names or titles below the objects.

Whereas, in the Incomplete Figures Task, the behavioral set restricting performance was closure, in the Circles task the individual is confronted with a closed figure. Yet the commonness of the circle stimulus causes many people to respond with sets that include only objects synonymous with the plain circular shape, i.e., ball, face, sun. The instructions ask for responses incorporating all four traits being assessed: fluency (see how many objects or pictures you can make); flexibility (make as many different pictures or objects as you can); originality (try to think of things that no one else will think of); and elaboration (put as many ideas as you can in each one and make them tell as complete and as interesting a story as you can). The time, however, is not adequate to give emphasis to all four kinds of thinking. Thus, individual preferences or response tendencies will come into play.

The scoring is essentially the same as that for the Incomplete Figures task. Fluency is again the number of relevant responses, although for the Circles task any repetitious responses are subtracted from this total. Flexibility is again obtained by counting the number of different categories into which the subject's responses can be classified. There are 60 categories listed in the manual, essentially the same as for Incomplete Figures. Originality is obtained by means of a scoring guide. 0 responses, 1 point responses, and 2 point responses are listed. Again the instructions call for all other responses not listed, showing creative strength to be assigned scoring weights of 2 points. Examples of 0 responses are: (a) apple, (b) door knob, (c) moon, (d) steering wheel. Examples of 1 point responses are: (a) ash tray, (b) ice cream cone, (c) lamps, lantern, lampshade, (d) records. 2 point responses listed include: (a) aquarium, aquarium with fish, (b) fruits (other than apple or orange), (c) locks, (d) rocket. These weights are based on tabulations of

responses by 588 subjects. The principles for scoring elaboration for the Circles task are the same as those which have been stated for the Incomplete Figures task. An example of elaboration of a stimulus labeled as apple would include: a stem, leaves, a rotten spot, a worm coming out of the apple, a pitted shape around the stem, a slight elevation on the side opposite the stem, etc. An additional point would be added for each of these ideas.

## Verbal tasks

### Product Improvement

This task requires the subject to formulate suggestions for improving an object with respect to its function as a toy. The stimulus selected for use was a stuffed toy dog. This was chosen from among the various stimuli experimented with by Torrance because it most successfully minimizes differences in the quality and nature of responses based on sex differences. For individual administration a stuffed toy dog is placed on the table in front of the child. For group administration, a sketch of a toy dog accompanies the sheets with instructions and blanks for the responses; and a stuffed toy dog is placed in view of all subjects who are told that they may come up and look at the dog or handle it if they so desire. The instructions given are:

> Try to think of the cleverest, most interesting and most unusual ways you can for changing this toy dog so that boys and girls will have more fun playing with it. Don't worry about how much it would cost -- just so it would make it more fun to play with.

When the test was individually administered, the instructions were concluded with:

> As fast as you think of ideas, tell them to me and I'll write them down for you.

When the test was group administered, the instructions were:

> In the spaces on this page and on the next one, list the cleverest, most interesting and unusual ways you can think of for improving this toy dog.

Torrance reports that this task has been one of his most dependable verbal tests. It is a complex test with a high degree of face validity. It permits subjects to consider ideas which they would not dare express in a more serious task. The nature of the task allows for scoring on the traits of fluency, flexibility, originality and elaboration. The fluency score is again determined by counting the number of relevant, separate responses or ideas given by the subject. An idea is relevant if it tells how the toy dog could be improved as a <u>plaything, something which will be fun</u>. The flexibility score is determined by the number of different principles or approaches used in responding to the task. An inclusive list of twenty general principles is given in the scoring manual. Examples of these principles are: (1) adaptation; change it to a cat, mouse, etc. Play uses other than as a toy dog; (2) Change color; (3) Division: cut into parts, etc. (4) Multiplication: Litter of dogs, pairs, triplets, etc. (5) Ear appeal: barks, makes noise when moved, sings a song, etc. (6) Subtraction: take off seal, take bow off, etc. Originality is determined from a table of responses giving scoring weights of 0, 1, or 2. Approximately 350 responses are listed, based on a tabulation of the responses of 594 subjects, from grades one through twelve. Rare responses

showing creative strength but not included in the list are to be scored "2". Examples of 0 responses are: (1) Bark, speak; (2) Change eye color; (3) Fuzzy or hairy; (4) Legs move mechanically; (5) Add Ribbon; (6) Make smaller; (7) Wrap it up. One point responses include: (1) All in one color; (2) Dish for food; (3) Float; (4) Neck moves; (5) Nose bright, shine, light up; (6) Able to swim. Two point responses include: (1) Adjustable in length; (2) Dish of water; (3) Fleas on it; (4) Mustache or beard; (5) Button nose; (6) Tail stands up; (7) Washable. Elaboration is determined by scoring the extent to which the idea is spelled out or elaborated by counting the details beyond that which are necessary to communicate the basic idea. Illustration of responses scored for elaboration are: Put a bone in its mouth (0); Have its eyes open and close (0); Give him whiskers that stretch when pulled upon* (1) (the asterick refers to the content that is an elaboration) Make it of candy. When the child tires of it, he could eat it* (1); Make a plaid* blanket for winter* (2); Put in an ink supply inside each foot* Wind up dog* and as it walks the dog will track up the floor* (3).

## Unusual Uses of Tin Cans

The Unusual Uses of Tin Cans task is a direct modification of Guilford's Brick Uses Test. After preliminary tryouts with a variety of stimuli, Torrance and his staff decided to substitute "tin cans" for "bricks". It was believed that children would be able to ideationally handle "tin cans" more easily than "bricks", since tin cans are more readily available to children for experimentation, manipulation, and playing. The instructions for the task are:

> Most people throw their empty tin cans away, but they have thousands of interesting and unusual uses.

> In the space below and on the next page, list as many of these interesting and unusual uses as you can think of. (For individualized administration this was changed to 'As you think of unusual uses for tin cans, tell them to me and I will write them down'). Do not limit yourself to any one size of can. You may use as many cans as you like. Do not limit yourself to the uses you have seen or heard about; think about as many possible new uses as you can.

It was recognized at the outset that "tin cans" create, in many individuals, rigid sets which are difficult for them to overcome. It is easy to define a tin can as a "container" and then to think of all the different things which can be put into tin cans, making it difficult to think of other types of response. Thus, the task is, in part, a test of ability to free one's mind of a well-established set.

In scoring, Fluency scores are determined by counting the number of different, relevant responses. The duplicate responses and responses which are irrelevant to the task are eliminated. The Flexibility score is determined by counting the number of different categories into which responses fall. The categories were determined in essentially the same manner as were the categories for the Incomplete Figures test, and the Circles task. The scoring weights for Originality were determined by a tabulation of the responses of 256 subjects, in the same manner as for the other tasks. Examples of 0 responses are: (1) Animal or bird waterer or feeder; (2) Kick the can game; (3) Seat, Chair, etc.; (4) Wedding car noise or decoration; Responses given 1 point include: (1) Apparel, hat, shoe, etc.; (2) Incinerator; (3) Shoe scraper; (4) Yarn Winder. 2 point responses include: (1) Burglar alarm, etc.; (2) Kaliedoscope (with rocks inside); (3) Sanitation, pottie, etc.; (4) Weapon. The rationale for scoring elaboration of the Tin Can Uses test is the same as for the Product Improvement test.

On the basis of observation made during the pilot study, the decision was made to administer the Non-Verbal tasks to all age levels in groups. This was done with groups no larger than twelve. At the younger ages it was necessary to help the students write the name or title of their picture after they told the examiner what it was. The verbal tasks were administered individually to the children at grades one, four and six. This was done as a result of the observation that a number of the students in the age range for the sixth grade in Prince Edward County were unable to express themselves in writing with any degree of ease. The eighth grade students were administered the verbal tasks in groups of twelve. All subjects took the nonverbal test prior to taking the verbal test.

The problem of reliability is always a sensitive one for newly developed instruments. The Minnesota Tests of Creative Thinking are rather new, having been developed since 1959, and frequent revisions in the first several years of their use have hampered reliability studies. However, Torrance does report some reliability figures and related issues (Torrance 1963c). He points out that in addition to their newness, several other problems make it difficult to show high reliability.

For one thing the subtests are often relatively short and this tends to reduce reliability. For other subtests like Product Improvement and Unusual Uses, there is only one item, or if one equates item to response for reliability purposes, the length becomes a dependent variable in itself. Secondly, and most crucially, a majority of the stability correlations available from test-retest situations are derived from experiments in which a deliberate effort has been made to encourage creativity during the interval. A third problem is that common to most psychological testing; that of lower reliability for tests

used with younger children. A final factor is simply that since these tests are quite unlike tests that most subjects have taken before, their initial performance may be affected by the novelty of the test. In the retest situation this effect is reduced, but this source of variation can serve to reduce reliability.

Torrance has collected limited reliability figures. Some of the reliability data has been gathered at the college level. Forty students were pretested, middle-tested, and post tested. After the first two weeks, the reliability figures for the Circles task were: Fluency (.92), Flexibility (.60), Originality (.63), Elaboration (.63). The figures for the two week gap between the middle test and the post test were: Fluency (.65), Flexibility (.62), Originality (.81), and Elaboration (.73). The reliability figures for the four week period from pretest to post-test were: Fluency (.47), Flexibility (.60), Originality (.57) and Elaboration (.65). These figures were obtained in situations where deliberate efforts were made to maximize creativity. As a part of the same experiment the students were given the Unusual Uses task and the resulting reliability was: for the first two weeks, Fluency (.61), Flexibility (.62), and Originality (.71); for the second two week period, Fluency (.75), Flexibility (.74), and Originality (.66); for the four week period, Fluency (.65), Flexibility (.71), Originality (.60).

In other work with college students, twenty-two were tested and then retested ten weeks later without an intervening effort to maximize creative expression. For the Circles task the reliabilities were: Fluency (.76), Flexibility (.63), Originality (.79), Elaboration (.67), and for the total score (obtained by summing the four sub-scores or scale scores) the reliability was (.75). Product Improvement reliability for the same study showed: (.69), (.64), (.61), (.75), and (.73) respectively, and for Unusual Uses: (.85), (.60), (.64), (.57), and (.68).

A study done with 101 ninth graders, where an effort was made to maximize creativity during the one week interval between pre-test and post-test, a reliability of (.69) was obtained for the total score on the Circles test. A reliability of (.80) for the total score on the Incomplete Figures test was obtained over a two week interval on tests administered to fifty-four ninth graders. Whether the subjects were exposed to any treatment effect during the two weeks was not stated.

In a study involving 100 to 150 students at each grade level from second through fifth grade reliabilities of (.34) to (.79) were obtained on total score figures from a Fall test-Spring retest situation. This was over a complete battery, which included several verbal tasks not included in the Abbreviated Form VII utilized in this study. Although this information is too general to clarify the matter of reliability, it is significant in that the experimenter found that reliabilities increased with age. These findings are in accord with those of the Bureau of Educational Research at the University of Minnesota, and indeed are reflective of conditions common to psychometric endeavors.

Another study using a battery of verbal tests, which included the Product Improvement and Unusual Uses tasks, plus several more, reported reliability coefficients of Fluency (.82), Flexibility (.78), and Originality (.59) and total score (.83). This study was done on 85 college students with a ten week interval during which attempts were made to maximize original thinking. Another study done with fifty fourth through ninth graders with a two week interval between testing situations reported a total score reliability of (.79) on a battery of verbal tests which included the two verbal tasks utilized in this study.

This smattering of reported data is inadequate for drawing firm conclusions relative to the adequacy of the reliability of the various tests of creativity developed by Torrance, and those specific tasks used in this study. At best it can be said that these tests have demonstrated a marginal reliability suitable for research purposes, although any conclusions derived from such research should be interpreted subject to the limitations of the reliability factors.

## Methods of Data Collection and Data Processing

The original plans for the administration of the "creativity" tests to the two population samples were for two people to do all the testing. These two were the major investigator and a fellow graduate student at Michigan State University. The latter individual had had a number of years of experience as an elementary teacher and worked exceptionally well with children in a one to one setting.

However, these plans, for various reasons, could not be put into operation. The entire Prince Edward County research project was temporarily suspended and thus this particular aspect of the study was placed in the same jeopardy. When negotiations were finally successful and the research plans reinstated, enough time had elapsed that the procedures had to be altered for the following reasons:

(a) Only three and one-half days were available for testing in Virginia necessitating more than two examiners.
(b) The timing of the Virginia trip was such that it was necessary to collect data from the Michigan sample at the same time. This was primarily a function of the proximity of the closing date of school for the

latter group.

As a result the actual administration of the creative test battery was done by the following personnel.

A. Michigan

(1) All but one of the group testing situations was administered by the above mentioned graduate student from Michigan State University. The remaining test was administered by the principal investigator. (Two days of school remained in the Michigan schools following the return from Virginia).

(2) The majority of the individual tests were administered by two undergraduate students from Michigan State University. Both examiners were given training in the administration of the test with children of the appropriate ages. This training took place at an elementary school in Okemos, Michigan, and at several Sunday Schools in the East Lansing area and was done under the supervision of the principal investigator. Because of absences it was necessary for the principal investigator to administer some individual tests to children of this sample following his return from Virginia. Because the Michigan sample, Grade Levels 1, 4 and 6 was drawn from two elementary schools, each of the examiners tested all three age levels.

B. Prince Edward County

(1) All of the group testing was administered by the principal investigator.

(2) All of the individual testing was done by two individuals on the substitute teacher roster of the Freedom Schools. These individuals were also trained,

under the supervision of the principal investigator, using subjects of the appropriate age levels from the classes of the Freedom Schools which were not included in the sample for the study. As in Jackson, Michigan, the sample for grade levels 1, 4 and 6 was drawn from two schools and each examiner worked in a given school with children from all three age levels.

Upon completing the collection of data, the protocols were given to a secretary who (a) scored the protocols by task, (b) randomly mixed the protocols from the different sample groups, both by county and age, and (c) assigned each subject a code number and wrote this on the protocol and then blacked out the subject's name. This procedure was done separately for each task so that the code number for any given subject was different for each of the four tasks. The protocols were then returned to the principal investigator for scoring.

All of the scoring was done by the principal investigator. A "score-rescore" procedure, with a lapse of two months, had been run on the pilot study data. The resulting "score-rescore or intra-rater reliabilities" were obtained:

| | |
|---|---|
| Incomplete Figures | .98 |
| Circles | .93 |
| Product Improvement | .98 |
| Uses | .97 |

The reliabilities by Scale were:

| | |
|---|---|
| Fluency | .99 |
| Flexibility | .99 |

| | |
|---|---|
| Originality | .94 |
| Elaboration | .87 |

Once the scoring had been completed the subjects were systematically coded based on county group, sex, and age. The four code sheets prepared by the secretary were then used to identify the subjects protocols and the scores were recorded on computer work sheets for analysis.

The computer analysis was used to:

(a) determine mean and standard deviation for subjects grouped by county, age, and sex. Thus there were a total of sixteen separate sets of means and standard deviations produced. Each set contained the statistics for thirty-five variables; each scale score and total score for each of the four tasks, non-verbal and verbal subtotals for each scale and a total verbal and non-verbal score, and a total score for each scale plus the grand total score. In addition, the computer was utilized to produce a correlational matrix of all variables. Three such matrixes were computed; one for the total sample and one each for the Michigan and Virginia samples. All three matrixes contained the thirty-five variables above, plus the variables of sex and age, and the total sample matrix also included the variable of county.

## Hypotheses

Given the rationale expressed in the preceding chapter, the traits of the two populations to be compared, and the nature of the assessment techniques used to generate the comparisons, the

following hypotheses are offered:

(1) Hypothesis: The performance scores of the Prince Edward County, Virginia, children will be significantly higher than the scores of the Jackson, Michigan children.

Because the nature of prevalent school objectives and methods tend, in numerous ways, to be suppressive, or at best non-supportive of many of those traits associated with creative thinking, it is predicted that across both sexes, all ages, and all scoring scales, the Virginia students, whose involvement in formal schooling has been less than typical, will perform significantly better on creative tasks, than the Michigan students, whose involvement, in terms of schooling, has been typical.

The assumptions being tested will also be examined as a sequence of four sub-hypotheses.

(a) Hypothesis: There will be no significant differences in the performance of the Virginia students and the Michigan students in the six year old group.

As has been indicated earlier in this chapter, the assumption is made that the two six year old groups will be essentially identical, both having experienced one year of formal schooling. This assumption is in turn predicated on two other assumptions. The first is that the one year of kindergarten experience for the Jackson, Michigan, students will not produce differences in performance on creative tasks. The former assumption also assumes the absence of socio-cultural differences between the two populations that might be determinants of criterion behavior.

To give the rationale for the remaining three sub-hypotheses it is necessary to refer to two points developed in the preceding

material. The first of these is the assumption that the first several years of schooling are the most crucial in determining the individual's ability to perform on tests of creativity. The second point is that of the differentiated school experience of the three pairs of age groups. This can best be summarized as: (1) Jackson, Michigan, fourth grade children had completed their fourth year of school involvement while the Prince Edward County, Virginia, children had completed their first year; (2) the sixth grade students in Michigan were completing their sixth year of school experience while the Virginia children had had school for one year, then were out four years and back in for one year, for a total of two years of experience; and (3) the eighth grade students in Michigan had completed their eighth year while the comparable Virginia students had had three years of schooling, then four years of no school and then an additional year of schooling, for a total of four years.

Thus, since the fourth grade and sixth grade students in Virginia had had significantly less formal school exposure during the first several school years, the following hypotheses are predicted:

(b) Hypothesis: The fourth grade (9 year old) Virginia students will do significantly better on the tests of creativity than will the fourth grade Michigan students.

(c) Hypothesis: The sixth grade (11 year old) Virginia students will do significantly better on the tests of creativity than will the sixth grade Michigan students.

Finally, even though there is a difference of four years of formal school experience between the 13 year old children, eighth graders, in Michigan and Virginia, both groups had had identical

involvement in terms of time, during the first three years of school. Therefore it is predicted:

(d) Hypothesis: There will be no significant difference in performance scores on the tests of creativity between the Virginia eighth grade students and the eighth grade Michigan students.

(2) Hypothesis: The factor of age will make significant differences in scores, with performance scores increasing with age.

It is reasonable to assume that the child's ability to perform on tasks of the nature employed by this study will increase with age. This is assumed not only for those traits or skills directly identified with creativity, but also for those factors which can facilitate creative expression e.g. verbal, ability, handling of pencils, etc.

(3) Hypothesis: There will be no difference in performance when the sample is compared by sex.

Although some of the literature has indicated sex differences in creative expression, these differences have tended to be inconsistent, and more related to specific types of tasks.

(4) Hypothesis: Scores on the non-verbal forms in comparison to the verbal forms will be significantly higher for the total population studied.

The battery of tests administered in this study is divided into two forms, verbal and nonverbal. Two factors would seem to support the idea of differential performance on the two forms. One of these factors is that of age. As a consequence of including students ranging from first grade to eighth grade, the samples

include many subjects for whom verbal development, by general developmental principles, is incomplete. Secondly, the particular characteristics of both groups place them in the currently popular rubric of culturally disadvantaged students. The diagnosis made of this group universally emphasizes a restricted verbal development.

If the first part of the rationale for hypothesis four is correct, then the following hypothesis is also warranted.

(5) Hypothesis: The differences between verbal and non-verbal performance will decrease significantly with each increase in age.

As a consequence of the differential school experience, it is also predicted that performance scores on the two different forms will be different for Michigan and Virginia students. This factor can best be tested with two sub-hypotheses that reflect the contribution of schooling to verbal development.

(6) (a) Hypothesis: The Virginia students performance scores on non-verbal tests of creativity will be significantly higher than the scores obtained by the Michigan students.

(b) Hypothesis: The Michigan students performance scores on verbal tests of creativity will be significantly higher than the scores obtained by the Virginia students.

If these two hypotheses are to be consistent with the first hypotheses, it would follow that the nonverbal differences in favor of the Virginia students will be greater than the verbal differences in favor of the Michigan students.

## Design

The general purpose of this study was to examine the effects of lack of formal schooling upon performance on tests of creativity. Samples from the two populations, Prince Edward County, Virginia and Jackson, Michigan, were stratified on the basis of sex and of grade, the grade factor having four levels - grade 1, grade 4, grade 6, and grade 8.

The performance scores on the four tasks of the creativity battery were analyzed in various combinations. Scores on the four tasks of tests were not used, as such, but rather fifteen scores for each individual were used in the analysis.

(a) One (1) total score. This score is the sum of the subjects total score on each of the four tasks.

(b) Four (4) Total Scale scores, Fluency, Flexibility, Originality, and Elaboration. The total scale score is the sum of the scores for each scale on each task.

(c) Two (2) Total Form Scores, Verbal and Non-Verbal. The total Form score is derived by summing the total task score for each of the two non-verbal tasks and the two verbal tasks.

(d) Eight (8) Form Scale Scores. These scores are the subtotal scale scores for each of the two forms, e.g. Verbal Fluency, Non-verbal Fluency, etc.

These scores and their derivations can be understood best by a replication of the form used for recording scores. Those scores underlined are the fifteen which are utilized in the analysis.

Table 1

Format used for recording individual scores and derivation of scores used in statistical analysis

| | Fluency | Flexibility | Originality | Elaboration | Total Score |
|---|---|---|---|---|---|
| Incomplete Figures | a | b | c | d | a+b+c+d |
| Circles | e | f | g | h | e+f+g+h |
| Non-Verbal Scale Scores | a+e | b+f | c+g | d+h | Total Non-Verbal Score a+b...+h |
| Product Improvement | i | j | k | l | i+j+k+l |
| Uses | m | n | o | p | m+n+o+p |
| Verbal Scale Scores | i+m | j+n | k+o | l+p | Total Verbal Score i+j...+p |
| Total Scale Scores | a+e+i+m | b+f+j+n | c+g+k+o | d+h+l+p | a+b+...+p |

As has been indicated in the description of the population on pages 25-27, the two groups studied were assumed to be comparable except for the differences in amount of formal schooling. The purpose of the analysis was to ascertain whether there were detectable differences between these two populations on the various creative performance criteria specified previously. An analysis of variance design was used to achieve these objectives.

The design is that of a five factor repeated measure design with two of the factors comprising the repeated measures. The independent factors of county, sex, and age are completely crossed as are the repeated measures of scale and form (resulting in a 2x2x4x4x2 design). All factors were seen as fixed. Since there were unequal, but not widely disparate, numbers in each cell, the analysis was performed using the harmonic mean for the number in each cell (see Winer, 1962, pp. 222-223). The use of the harmonic mean allows one to confidently utilize the analysis of variance procedure designed for cells with equal "n's". This technique does, however, somewhat lower the power of the analysis procedure with consequent conservative significance tests. However, as will be further explained, the attempt was deliberately made to make the necessary statistical adjustments conservative to compensate for the marginal reliability of the assessment techniques.

Table 2 illustrates the analysis of variance model employed.

While there are problems associated with using an analysis of variance model with these data, these problems are largely a function of the treatment being carried out prior to and independent of this study, and would present equal difficulties in attempting to apply any other statistical technique, e.g., multiple "t" tests or separate analysis of variance for each of the repeated measures. The basic problem is one of not being able to place full confidence in the obtained significance levels. However, the previously mentioned multiple techniques not only have all the difficulties of an analysis of variance but also introduce a degree of dependence among the various significance tests which would further distort obtained probability levels.

The application of the analysis of variance technique when repeated measures of subjects are employed raises valid questions

Table 2

Analysis of variance model employed

| County | Sex | Age | Scales | Fluency | | Flexibility | | Originality | | Elaboration | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | Forms | NV | V | NV | V | NV | V | NV | V |
| | | | Individuals | | | | | | | | |
| Virginia | Boys | 6 | | | | | | | | | |
| | | 9 | | | | | | | | | |
| | | 11 | | | | | | | | | |
| | | 13 | | | | | | | | | |
| | Girls | 6 | | | | | | | | | |
| | | 9 | | | | | | | | | |
| | | 11 | | | | | | | | | |
| | | 13 | | | | | | | | | |
| Michigan | Boys | 6 | | | | | | | | | |
| | | 9 | | | | | | | | | |
| | | 11 | | | | | | | | | |
| | | 13 | | | | | | | | | |
| | Girls | 6 | | | | | | | | | |
| | | 9 | | | | | | | | | |
| | | 11 | | | | | | | | | |
| | | 13 | | | | | | | | | |

NV = Nonverbal
V = Verbal

concerning certain assumptions about the "F" ratio actually following an "F" distribution. Primarily it is a matter of the assumption that sources of variance within subjects (subjects within groups) are homogeneous. Additional assumptions about the patterns of elements must be met for the sampling distributions of the "F" ratio for within subjects effects to be the "F" distribution with the usual degrees of freedom. A method of adjusting the analysis of repeated measures to avoid these assumptions is provided by Greenhouse and Geisser (Winer, 1962). The approximation procedure given by these writers errs in the direction of making the critical value somewhat larger than should be the case, and thus represents another conservative adjustment. In the Greenhouse and Geisser procedure the "F" ratio originally obtained is used but the degrees

of freedom used in finding the critical values are adjusted. In principle, the degrees of freedom for the numerator and the demoninator of the "F" distributions required in the usual tests (as given by the expected values of the mean square obtained) are divided by the degrees of freedom for the factor on which there are repeated measures.

To test certain of the hypotheses it was necessary to employ a multiple comparison technique to supplement the basic analysis of variance design. The technique selected was the Neuman-Keuls procedure. (Winer, 1962) This is a method for testing the significance of the differences between all possible pairs of means. The harmonic mean of 10.9 was again used in this analysis. This technique was selected from other alternative methods because it represents a compromise between the approaches suggested by Duncan and the more conservative approach of Scheffe.

## Chapter 4 - Primary Analysis of the Data

The data derived from this study will be presented in two chapters. Chapter 4 summarizes the results of the reliability study run on two of four "Minnesota Tests of Creativity" utilized for this project, and presents the results of the analysis of variance and associated methods used to test the hypotheses formulated in Chapter 3. Chapter 5 presents various analyses of the data not directly relevelant to the aims of the study but which in various ways facilitate an interpretation of the data.

### Results of Reliability Study

As has been indicated earlier, reliability figures are relatively scarce for the "Minnesota Tests of Creativity." Thus, when circumstances made it possible to repeat the administration of two of the tests to one of the population samples involved in this study, it became possible to include a partial reliability study. Unfortunately, it is only partial. Only the Prince Edward County children were retested and some had left the

vicinity or were absent on the days of test administration. The partial designation is also a consequence of retesting only two of the four tests of the battery, the nonverbal tasks "Incomplete Figures" and "Circles." Despite these limitations, it is possible to make limited inferences concerning the general reliability of the MINNESOTA TEST OF CREATIVE THINKING, ABBREVIATED FORM VII. (Torrance, 1963) These results and related materials are presented in Appendix A.

The reliability study utilized the test-retest procedure on identical test forms with approximately a one year interval between testing sessions. Thus the procedures differed from those used in studies from which Torrance derived his reliability figures in two major respects. First, the time interval between testing sessions was much longer, one year as opposed to several weeks. It is significant to note that the one study reported by Torrance with Fall, Spring test-retest schedule showed the widest range in reliabilities. Second, there was no intervening treatment effect to encourage or maximize creativity in the study reported herein, whereas this was a pervasive characteristic of the studies used by Torrance in reporting reliability.

The range of reliabilities using scale scores on the individual tests was .17 to .65. Total non-verbal scale score reliabilities were as follows: Fluency, .60; Flexibility, .67; Originality, .48; and Elaboration, .68. The reliabilities for the two tests were Figure Completion .53 and Circles .66, and the total non verbal score had a reliability of .69. The table in Appendix A breaks these figures down by sex and by grade level with a much wider range than reported for the total group. The sample size for the total group was 77.

These inferences with respect to the total battery are enhanced by the fact that the reliability studies reported in the literature had consistently found lower coefficients of reliability for the two non-verbal tasks in comparison to the two verbal tasks. Thus, if would seem reasonable to assume that the reliability figures shown in Appendix A represent the lower limits of reliability for the total battery. The obtained reliability data reflects the general trend of an increase in reliability with age. This is the pattern consistently reported in other reliability studies. This phenomenon is by no means limited to efforts to assess creative behavior.

Although the breakdown of the reliability data by sex and age level reveals pockets of extremely low, even negative coefficients of reliability, the same general conclusion can be drawn from the collective reliability figures as was drawn from a review of the reliability studies reported by Torrance (1963e). The tests of creative thinking used for this study show marginal reliability, adequate for use in research purposes if subsequent conclusions are interpreted in light of the limitations imposed by assessment procedures of relatively low reliability. This reliability study, like the others reported, demonstrate the need for improvement in these and similar procedures for assessing creative behavior.

## Testing the Hypotheses

Table 3 contains the group means and standard deviation on the fifteen creativity scores described on page 51 for each age group, separated by sex, in the two sample groups.

Table 4 contains the results of the analysis of variance.

Table 3

Means and standard deviations for fifteen creative performance scores for the 16 sample groups (County x Sex x Age) total n = 175

In each cell the top figure is the mean, the lower figure, the standard deviation

| | | | Non-Verbal Fluency | Non-Verbal Flexibility | Non-Verbal Originality | Non-Verbal Elaboration | Non-Verbal Total Score | Verbal Fluency | Verbal Flexibility |
|---|---|---|---|---|---|---|---|---|---|
| VIRGINIA | GIRLS | 6(N=10) | 20.0 | 12.2 | 9.0 | 11.1 | 52.3 | 13.4 | 7.0 |
| | | | 3.20 | 4.18 | 3.74 | 5.22 | 12.04 | 6.02 | 3.37 |
| | | 9(N=12) | 19.8 | 15.0 | 10.5 | 15.1 | 60.3 | 9.3 | 6.5 |
| | | | 5.10 | 4.02 | 4.25 | 9.50 | 19.17 | 4.96 | 3.00 |
| | | 11(N=12) | 18.8 | 14.8 | 8.1 | 15.8 | 57.4 | 12.6 | 6.8 |
| | | | 4.54 | 2.98 | 4.25 | 7.70 | 16.32 | 4.83 | 2.67 |
| | | 13(N=11) | 23.0 | 17.5 | 13.0 | 31.8 | 85.4 | 17.0 | 9.9 |
| | | | 4.10 | 3.47 | 4.75 | 12.20 | 21.03 | 6.03 | 3.86 |
| | BOYS | 6(N=10) | 20.0 | 13.6 | 11.3 | 8.9 | 53.8 | 13.4 | 8.1 |
| | | | 4.59 | 4.57 | 4.11 | 5.17 | 14.35 | 5.89 | 2.92 |
| | | 9(N=10) | 22.2 | 17.1 | 9.7 | 11.6 | 60.6 | 11.1 | 5.7 |
| | | | 7.00 | 4.79 | 3.47 | 7.89 | 20.25 | 7.00 | 1.94 |
| | | 11(N=10) | 20.3 | 14.9 | 11.9 | 11.5 | 58.6 | 9.9 | 5.7 |
| | | | 6.63 | 3.18 | 4.01 | 6.08 | 13.18 | 5.22 | 2.41 |
| | | 13(N=10) | 20.4 | 12.0 | 8.7 | 22.4 | 63.5 | 9.0 | 6.5 |
| | | | 6.93 | 4.35 | 5.25 | 12.23 | 20.23 | 4.67 | 3.41 |
| MICHIGAN | GIRLS | 6(N=11) | 13.5 | 9.6 | 10.5 | 7.2 | 40.9 | 20.5 | 10.3 |
| | | | 4.78 | 3.07 | 5.11 | 7.22 | 16.06 | 6.90 | 5.90 |
| | | 9(N=12) | 17.0 | 11.1 | 11.9 | 23.5 | 63.5 | 13.4 | 7.4 |
| | | | 5.38 | 3.85 | 6.14 | 14.03 | 22.23 | 6.67 | 3.18 |
| | | 11(N=12) | 20.3 | 14.3 | 12.8 | 28.0 | 75.3 | 11.9 | 8.6 |
| | | | 7.44 | 3.91 | 6.00 | 9.98 | 23.60 | 5.04 | 2.87 |
| | | 13(N=13) | 14.0 | 11.0 | 8.5 | 16.2 | 49.6 | 18.4 | 10.2 |
| | | | 2.24 | 2.24 | 2.90 | 7.15 | 10.46 | 5.42 | 3.02 |
| | BOYS | 6(N=10) | 13.4 | 10.5 | 8.0 | 5.2 | 37.1 | 13.8 | 7.8 |
| | | | 5.82 | 4.06 | 3.97 | 3.22 | 14.65 | 7.87 | 3.82 |
| | | 9(N=11) | 13.6 | 10.2 | 10.5 | 14.1 | 48.4 | 13.2 | 8.2 |
| | | | 3.67 | 3.60 | 2.81 | 4.64 | 9.73 | 8.04 | 4.12 |
| | | 11(N=12) | 21.0 | 13.8 | 11.7 | 24.8 | 71.3 | 14.5 | 9.3 |
| | | | 5.66 | 3.54 | 3.63 | 13.27 | 16.94 | 7.95 | 3.98 |
| | | 13(N= 9) | 26.4 | 17.7 | 13.1 | 30.8 | 88.0 | 19.1 | 13.4 |
| | | | 6.15 | 2.78 | 5.84 | 17.18 | 24.78 | 4.08 | 3.78 |

| Verbal Originality | Verbal Elaboration | Verbal Total Score | Total Fluency Score | Total Flexibility Score | Total Originality Score | Total Elaboration Score | Total Score |
|---|---|---|---|---|---|---|---|
| 6.9 | 4.9 | 32.2 | 33.4 | 19.2 | 15.9 | 16.0 | 84.5 |
| 4,58 | 2,96 | 14,05 | 8,49 | 5,75 | 7,43 | 5,83 | 22,5 |
| 2.9 | 3.0 | 21.7 | 29.0 | 21.5 | 13.4 | 18.1 | 82.0 |
| 2,61 | 1,91 | 10,97 | 7,95 | 5,76 | 5,88 | 10,00 | 25,74 |
| 3.0 | 4.2 | 26.5 | 31.3 | 21.6 | 11.1 | 19.9 | 83.9 |
| 2,13 | 2,04 | 9,05 | 6,95 | 4,87 | 5,25 | 8,56 | 19,75 |
| 8.7 | 5.1 | 40.7 | 40.0 | 27.5 | 21.7 | 36.9 | 126.1 |
| 5,85 | 3,73 | 17,01 | 8,61 | 5,99 | 8,70 | 13,34 | 31,33 |
| 7.2 | 6.6 | 35.0 | 33.4 | 21.7 | 18.2 | 15.5 | 88.8 |
| 3,33 | 3,89 | 14,61 | 7,93 | 5,93 | 5,07 | 6,45 | 21,40 |
| 5.0 | 2.4 | 24.4 | 33.3 | 22.8 | 14.7 | 14.0 | 85.0 |
| 2,67 | 2,01 | 11,27 | 9,27 | 5,53 | 4,14 | 7,92 | 20,29 |
| 2.3 | 1.8 | 19.7 | 30.2 | 20.6 | 14.2 | 13.3 | 78.3 |
| 1,49 | 1,81 | 9,15 | 8,82 | 4,03 | 4,49 | 7,38 | 18,74 |
| 4.6 | 2.5 | 22.6 | 29.4 | 18.5 | 13.3 | 24.9 | 86.1 |
| 4,48 | 2,51 | 13,60 | 8,60 | 6,38 | 8,17 | 13,74 | 31,28 |
| 13.5 | 3.7 | 47.9 | 34.0 | 19.9 | 24.0 | 10.9 | 88.8 |
| 10,87 | 4,54 | 23,73 | 8,21 | 6,19 | 11,76 | 10,61 | 30,81 |
| 8.1 | 4.4 | 33.3 | 30.4 | 18.5 | 20.0 | 27.9 | 96.8 |
| 4,01 | 6,24 | 17,02 | 10,10 | 5,74 | 8,71 | 17,45 | 34,46 |
| 6.3 | 4.0 | 30.8 | 32.3 | 22.8 | 19.0 | 32.0 | 106.1 |
| 3,62 | 4,02 | 12,11 | 8,13 | 5,22 | 7,93 | 10,93 | 25,36 |
| 11.3 | 12.7 | 52.5 | 32.3 | 21.2 | 19.8 | 28.8 | 102.2 |
| 7,45 | 5,47 | 13,13 | 5,36 | 3,83 | 7,45 | 10,15 | 18,44 |
| 8.5 | 3.4 | 31.7 | 26.3 | 17.4 | 16.5 | 8.6 | 68.8 |
| 9,88 | 5,70 | 23,60 | 9,76 | 6,81 | 11,61 | 7,57 | 31,29 |
| 11.1 | 3.3 | 35.7 | 26.8 | 18.4 | 21.5 | 17.4 | 84.1 |
| 11,42 | 2,97 | 24,32 | 10,52 | 6,19 | 12,63 | 4,99 | 31,57 |
| 8.6 | 5.8 | 38.1 | 35.5 | 23.2 | 20.3 | 30.6 | 109.4 |
| 8,30 | 5,77 | 22,67 | 9,86 | 4,80 | 9,69 | 15,66 | 30,03 |
| 15.4 | 4.1 | 52.1 | 45.6 | 31.1 | 28.6 | 34.9 | 140.1 |
| 2,07 | 1,83 | 10,95 | 7,47 | 5,69 | 6,19 | 18,63 | 33,55 |

Table 4

Results of analysis of variance

| Sources of Variation | d.f. | s.s | m.s. | demoninator & d.f. in f test | F and significance level | |
|---|---|---|---|---|---|---|
| Between Subjects | | | | | | |
| C(County) | 1 | 128.16 | 128.16 | (a) 1,159 | 1.39( .25 )' | |
| Q(Sex) | 1 | 446.25 | 446.25 | (a) 1,159 | 4.84( .05 ) | |
| A(Age) | 3 | 2987.70 | 995.90 | (a) 3,159 | 10.81( .001) | |
| CxQ | 1 | 201.8 | 201.8 | (a) 1,159 | 2.19( .25 )' | |
| CxA | 3 | 676.79 | 225.63 | (a) 3,159 | 2.45( .10 )' | |
| QxA | 3 | 44.49 | 14.83 | (a) 3,159 | .16(>.25 )' | |
| CxQxA | 3 | 2140.31 | 713.44 | (a) 3,159 | 7.74( .001) | |
| (a)1(CQA) | 159 | 14647.38 | 92.12 | | Adjusted d.f. and significance level | |
| Within Subjects | | | | | | |
| S(Scales) | 3 | 10348.88 | 3449.63 | (b) 3,477 | 6.89( .001) | 1,159 .01 |
| SxC | 3 | 236.40 | 78.80 | (b) 3,477 | 1.57( .25 )' | |
| SxQ | 3 | 481.34 | 160.45 | (b) 3,477 | 3.20( .05 ) | 1,159 .10' |
| SxA | 9 | 2150.63 | 238.96 | (b) 9,477 | 4.77( .001) | 3,159 .005 |
| SxCxQ | 3 | 21.12 | 7.04 | (b) 3,477 | .14(>.25 )' | |
| SxCxA | 9 | 994.44 | 110.49 | (b) 9,477 | 2.21( .025) | 3,159 .10' |
| SxQxA | 9 | 193.86 | 21.54 | (b) 9,477 | .43(>.25 )' | |
| SxCxQxA | 9 | 172.70 | 19.19 | (b) 9,477 | .38(>.25 )' | |
| (b)1(CQA)xS | 477 | 23898.78 | 50.10 | | | |
| F(Form) | 1 | 14960.41 | 14960.41 | (c) 1,159 | 10.31( .005) | 1,159 .005 |
| FxC | 1 | 473.78 | 473.78 | (c) 1,159 | 3.26( .10 )' | |
| FxQ | 1 | 364.13 | 364.13 | (c) 1,159 | 2.51( .25 )' | |
| FxA | 3 | 2399.39 | 799.80 | (c) 3,159 | 5.51( .005) | 3,159 .005 |
| FxCxQ | 1 | 34.43 | 34.43 | (c) 1,159 | .24(>.25 )' | |
| FxCxA | 3 | 770.30 | 256.77 | (c) 3,159 | 1.77( .25 )' | |
| FxQxA | 3 | 518.93 | 172.98 | (c) 3,159 | 1.19(>.25 )' | |
| FxCxQxA | 3 | 844.12 | 281.37 | (c) 3,159 | 1.94( .25 )' | |
| (c)1(CQA)xF | 159 | 23074.09 | 145.12 | | | |
| SxF | 3 | 4976.24 | 1658.75 | (d) 3,477 | 19.29( .001) | 1,159 .001 |
| SxFxC | 3 | 1067.16 | 355.72 | (d) 3,477 | 4.13( .01 ) | 1,159 .05 |
| SxFxQ | 3 | 477.74 | 159.25 | (d) 3,477 | 1.85( .25 )' | |
| SxFxA | 9 | 1748.73 | 194.30 | (d) 9,477 | 2.26( .025) | 3,159 .10' |

| | | | | | | |
|---|---|---|---|---|---|---|
| SxFxCxQ | 3 | 116.42 | 38.81 | (d) 3,477 | .45(>.25 )' | |
| SxFxCxA | 9 | 804.29 | 89.37 | (d) 9,477 | 1.04(>.25 )' | |
| SxFxQxA | 9 | 438.97 | 48.77 | (d) 9,477 | .57(>.25 )' | |
| SxFxCxQxA | 9 | 538.94 | 59.88 | (d) 9,477 | .70(>.25 )' | |
| (d)1(CQA)xSxF | 477 | 41015.49 | 85.99 | | | |

CODE: C - County (Prince Edward County, Virginia - Jackson, Michigan)
Q - Sex
A - Age (6, 9, 11, 13)
S - Scales (Fluency, Flexibility, Originality, Elaboration)
F - Form (nonverbal - verbal)
s.s. - Sum of squares
d.f. - Degrees of freedom
m.s. - Expected mean sum of squares

(a), (b), (c), (d) respective error terms (m.s.) for each of the four subsections of the table.

Those significance levels with ' (p.>.10) are assumed to be insignificant.

As can be noted in Table 4, where the main effect of Scales is introduced and throughout the remainder of the table, an additional column has been added labeled "Adjusted D.F. and Significance Level." This represents the application of the Greenhouse and Geisser technique (see page 54). When the original "F" tests result in a level of significance of the range .001-.05, this adjustment technique will be employed and only those adjusted significance levels which remain within the above range will be treated as significant.

The first section of the ANOVA table shows that the only main effect which does not contribute to the obtained differences is that of County. Over all Scales, Forms, Ages and Sex, there is no significant difference between that group of children from Virginia, whose time or amount of involvement in formal schooling has been restricted to varying degrees, and that group from Michigan, whose time or amount of school involvement has been typical for American public education. This result would clearly seem to reject

Hypothesis No. 1: "The mean performance scores of the Prince Edward County, Virginia, children will be significantly higher than the mean scores of the Jackson, Michigan children."

The four sub-hypotheses of the first hypothesis had predicted the following:

a. $\bar{X}v_6 = \bar{X}m_6$

b. $\bar{X}v_9 > \bar{X}m_9$

c. $\bar{X}v_{11} > \bar{X}m_{11}$

d. $\bar{X}v_{13} = \bar{X}m_{13}$

The analysis of the County x Age interaction represents the closest approximation of testing these hypotheses within the analysis of variance design. However, its results only tell us that the differences between age groups are not significantly different in the Virginia sample or the Michigan sample. The best means of testing this hypothesis is the Neuman-Keuls procedure. (See page 54). This analysis is presented in Table 5.

Given the results in Table 5, the following statement can be made: Hypothesis No. 1a: "$\bar{X}v_6 = \bar{X}m_6$." The results show that the mean score for Virginia children, age 6, is higher than that of their age mates in Michigan and that the difference is statistically significant at the .05 level. Hypothesis No. 1b: "$\bar{X}v_9 > \bar{X}m_9$." The results show a difference that is not statistically significant. However, the difference is in the predicted direction. Hypothesis No. 1c: "$\bar{X}v_{11} > \bar{X}m_{11}$." The analysis of the data shows significant differences at the .01 level but in the opposite direction of that predicted. Hypothesis No. 1d: "$\bar{X}v_{13} = \bar{X}m_{13}$." The analysis of the data shows no significant difference between the mean performance scores of these two subgroups. Thus, to summarize, the first three hypotheses were rejected and only sub-hypothesis No. 1d predicting equal performance

Table 5

Analysis of differences of pairs of means for the eight groups (Age x County) using the Neuman-Keuls procedure

| Age x County | $M_6$ | $V_{11}$ | $M_9$ | $V_6$ | $V_9$ | $M_{11}$ | $V_{13}$ | $M_{13}$ |
|---|---|---|---|---|---|---|---|---|
| | 78.9 | 84.0 | 86.3 | 87.6 | 87.6 | 105.1 | 109.4 | 117.8 |
| $M_6$ | | 5.1 | 7.4 | 8.7 | 8.7 | 26.2 | 30.5 | 38.9 |
| $V_{11}$ | | | 2.3 | 3.6 | 3.6 | 21.1 | 25.4 | 33.8 |
| $M_9$ | | | | 1.3 | 1.3 | 18.8 | 23.1 | 21.5 |
| $V_6$ | | | | | 0 | 17.5 | 21.8 | 30.2 |
| $V_9$ | | | | | | 17.5 | 21.8 | 30.2 |
| $M_{11}$ | | | | | | | 4.3 | 12.7 |
| $V_{13}$ | | | | | | | | 8.4 |

| $S_{\overline{A}}$=2.05 | r=2 | r=3 | r=4 | r=5 | r=6 | r=7 | r=8 |
|---|---|---|---|---|---|---|---|
| $\underline{9}.95(r,159)$ | 2.77 | 3.31 | 3.63 | 3.86 | 4.03 | 4.17 | 4.29 |
| $\underline{9}.99(r,159)$ | 3.64 | 4.12 | 4.40 | 4.60 | 4.76 | 4.88 | 4.99 |
| $S_{\overline{A}}\underline{9}.95(r,159)$ | 5.68 | 6.79 | 7.44 | 7.91 | 8.26 | 8.55 | 8.79 |
| $S_{\overline{A}}\underline{9}.98(r,159)$ | 7.46 | 8.45 | 9.02 | 9.43 | 9.76 | 10.00 | 10.23 |

| | $M_6$ | $V_{11}$ | $M_9$ | $V_6$ | $V_9$ | $M_{11}$ | $V_{13}$ | $M_{13}$ |
|---|---|---|---|---|---|---|---|---|
| $M_6$ | | n.s. | .05 | .05 | .05 | .01 | .01 | .01 |
| $V_{11}$ | | | n.s. | n.s. | n.s. | .01 | .01 | .01 |
| $M_9$ | | | | n.s. | n.s. | .01 | .01 | .01 |
| $V_6$ | | | | | n.s. | .01 | .01 | .01 |
| $V_9$ | | | | | | .01 | .01 | .01 |
| $M_{11}$ | | | | | | | n.s. | .01 |
| $V_{13}$ | | | | | | | | n.s. |

CODE: $M_6$ = Michigan 6 yr. olds
$M_9$ = Michigan 9 yr. olds
$M_{11}$ = Michigan 11 yr. olds
$M_{13}$ = Michigan 13 yr. olds
$V_6$ = Virginia 6 yr. olds
$V_9$ = Virginia 9 yr. olds
$V_{11}$ = Virginia 11 yr. olds
$V_{13}$ = Virginia 13 yr. olds

$S_{\overline{A}}$ = Standard error for the mean of the groups representing levels of factor A (age)

$\underline{9}.95(r,159)$ = Tabled values of the Studentized Range Statistics for $(1-\alpha) = .95$

$\underline{9}.99(r,159)$ = Tabled values of the Studentized Range Statistics for $(1-\alpha) = .99$

$S_{\bar{A}}\,q.95(r,159)$ = Critical values for an ordered difference between two means, r steps apart

$S_{\bar{A}}\,q.99(r,159)$ = Critical values for an ordered difference between two means, r steps apart

by the two samples at the eighth grade level was accepted. Although the performances were not equal, the difference was not statistically significant. These results will be discussed in Chapter 6.

The main effect of age across both counties, sex and form, in all scales, was significant at the .001 level. The result is as predicted in Hypothesis No. 2: "The factor of age will make significant differences in the scores." To further identify the cause of this difference an analysis of the age group means by the Neuman-Keuls procedure is employed. This analysis is shown in Table 6 for the main effect of age. The age groups show a consistent increase in mean performance scores as would be expected. This analysis indicates that a major source of the significant difference of the main effect is obtained from the disproportionate gain for the oldest age group, "$A_{13}$."

Thus, the difference in performance in scores by age is as predicted; and furthermore, the increase at each successive age level is as suggested in the rationale for Hypothesis No. 2 (p. 49). However, the data fail to fully support this hypothesis since the mean score differences are not significant between grade levels 1 and 4.

The main effect of Sex across both counties, at all age levels, for all scales and both forms is significant at the .05 level. The mean total score for the eight groups of boys 2 ("County") x 4 ("Age")

Table 6

**Analysis of differences of pairs of mean scores for subjects grouped by age - using the Neuman-Keuls procedure**

| Ages | $a_6$ | $a_9$ | $a_{11}$ | $a_{13}$ |
|---|---|---|---|---|
| Ordered Means | 83.23 | 86.93 | 94.53 | 113.58 |

| Differences Between Pairs | | | | |
|---|---|---|---|---|
| | $a_6$ | 3.70 | 11.30 | 30.35 |
| | $a_9$ | | 7.60 | 26.65 |
| | $a_{11}$ | | | 19.05 |

| $S_{\bar{A}}$ = 1.45 | r=2 | r=3 | r=4 |
|---|---|---|---|
| $\underline{9}.95(r,159)$ | 2.77 | 3.31 | 3.63 |
| $\underline{9}.99(r,159)$ | 3.64 | 4.12 | 4.40 |
| $S_{\bar{A}}\ \underline{9}.95(r,159)$ | 4.02 | 4.80 | 5.26 |
| $S_{\bar{A}}\ \underline{9}.99(r,159)$ | 5.28 | 5.97 | 6.38 |

| Significance of Differences | | | | |
|---|---|---|---|---|
| | $a_6$ | $a_9$ | $a_{11}$ | $a_{13}$ |
| | $a_6$ | n.s. | .01 | .01 |
| | $a_9$ | | .01 | .01 |
| | $a_{11}$ | | | .01 |

CODE: $a_6$ = age level 6 $a_{11}$ = age level 11
$a_9$ = age level 9 $a_{13}$ = age level 13
$S_{\bar{A}}$ = Standard error for the mean of the groups representing levels of factor A (age)
$\underline{9}.95(r,159)$ = Tabled values of the Studentized Range Statistics for $(1-\alpha) = .95$
$\underline{9}.99(r,159)$ = Tabled values of the Studentized Range Statistics for $(1-\alpha) = .99$
$S_{\bar{A}}\ \underline{9}.95(r,159)$ = Critical values for an ordered difference between two means, r steps apart
$S_{\bar{A}}\ \underline{9}.99(r,159)$ = Critical values for an ordered difference between two means, r steps apart

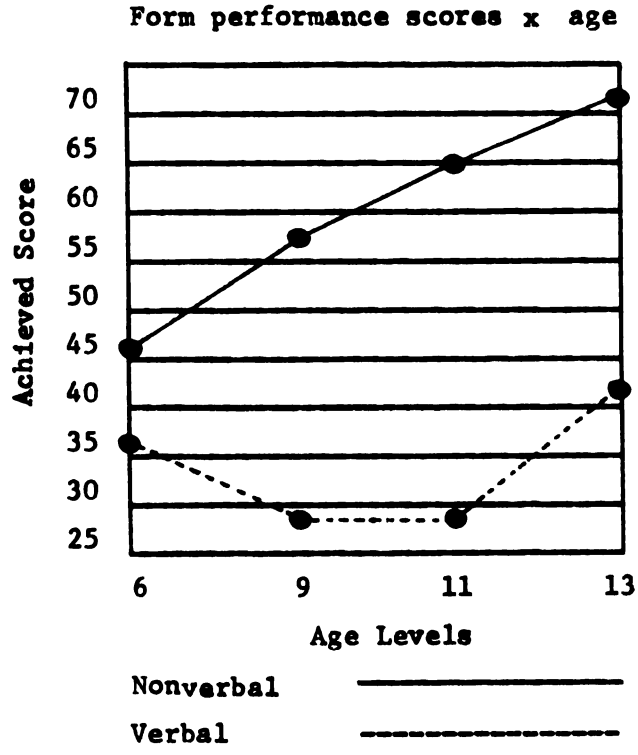
is 90.04, while for girls, it is 98.9. Thus, in terms of the main effect of sex, it can be stated that girls achieved higher scores than boys. The achieved results ran counter to the results predicted. Hypothesis No. 3 predicted: "There will be no significant difference in performance when compared by sex."

The main effect of Form was significant at the .005 level. The battery of four creative tasks, or tests, included two nonverbal tasks and two verbal tasks. Thus, the scores on the scales of Fluency, Flexibility, Originality and Elaboration can be dichotomized according to form, i.e., Verbal Flexibility, Nonverbal Flexibility, etc. The means for the 16 group means were 60.38 on the nonverbal and 34.18 for the verbal. Thus, Hypothesis No. 4 (The scores for the total study of populations will be significantly higher on the nonverbal forms in comparison to the verbal forms) is accepted as predicted.

Hypothesis No. 5 predicted that: "The differences between verbal and nonverbal performance will decrease significantly with an increase in age." The only other effect significant in this subsection of the analysis model was, indeed, the Form x Age interaction. This effect is illustrated by the graph in Figure 1.

This graph seriously challenges the validity of the assumption regarding age that was made in the rationale for Hypothesis No. 6. The trends reflected in the graph show an increase in the difference between verbal and nonverbal responses with age. The nonverbal scores show a consistent and steady increase with age. The nonverbal scores show a decrease in performance scores from age 6 to age 9, a plateau from age 9 to age 11, and a relatively sharp increase for age 13. This phenomenon raises some interesting and significant implications which can be more appropriately discussed in the following chapter. At this point the pertinent

Figure 1




observation is that the sources of significance for the Form x Age interaction is a consequence of the fact that nonverbal scores followed a standard increase in age while the verbal scores showed the decrease-plateau-increase pattern illustrated in Figure 1. Thus, Hypothesis No. 6 is rejected.

The interaction effect of Form x County was insignificant on the non-adjusted "F" test. However, because this interaction is the focus of Hypothesis No. 6, a further elaboration of this effect will be attempted. A breakdown of the group means shows the following:

| | Verbal | Nonverbal |
|---|---|---|
| Prince Edward County | 30.64 | 61.5 |
| Jackson, Michigan | 37.71 | 59.26 |

Thus, the near significance of the Form x County interaction effect can be primarily attributed to the relatively higher scores of the Michigan students on the verbal scales as compared with the verbal scores of the Virginia students. Hypothesis No. 6 is actually two sub-hypotheses: (a) The Virginia students performance on nonverbal tests of creativity will be significantly higher than the score obtained by the Michigan students and (b) the Michigan students mean performance scores on verbal tests of creativity will be significantly higher than the mean scores obtained by the Virginia students.

These hypotheses can best be tested using the Neuman-Keuls technique. This is shown in Table 7.

The results in Table 7 show that sub-hypothesis 6(a) Virginia Nonverbal Scores will be significantly higher than Michigan Nonverbal Scores is rejected, although the differences are in the right direction. Sub-hypothesis 6(b) is accepted. The Michigan Verbal scores were significantly higher than the Virginia Verbal Scores at the .01 level of significance. It should also be noted that Hypothesis No. 6 was accompanied by the suggestion that the nonverbal differences in favor of the Virginia students would be greater than the verbal differences in favor of the Michigan students. This is implied to maintain consistency with Hypothesis No. 1 which had predicted that the Virginia students would do better over all forms than the Michigan students. Since Hypothesis No. 1 was rejected, it is not too surprising that the Michigan students superiority on verbal performance was of a larger magnitude than the superiority of the Virginia students on nonverbal performance.

Figure 2 further illustrates the comparative form scores when the two counties are plotted separately.

Table 7

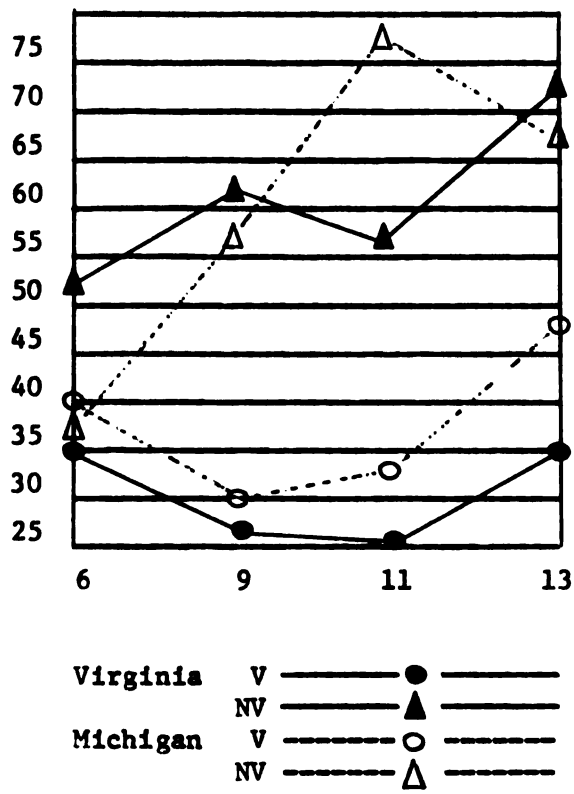
Neuman-Keuls test for significance of difference in verbal and nonverbal scores by county

| County x Form | V-V | M-V | M-NV | V-NV |
|---|---|---|---|---|
| Ordered Means | 30.64 | 37.71 | 59.26 | 61.5 |
| V-V Differences $r_1$ | | 7.07 | 28.62 | 30.06 |
| M-V $r_2$ | | | 21.55 | 23.7 |
| M-NV $r_3$ | | | | 2.24 |
| $S_{\bar{F}}$ = 1.29 | | | | |
| $\underline{9}$ = .95 (r,159) | | 3.77 | 3.31 | 3.63 |
| $\underline{9}$ = .99 (r,159) | | 3.64 | 4.12 | 4.40 |
| $S_{\bar{F}} = \underline{9}.95$ (r,159) | | 3.57 | 4.27 | 4.68 |
| $S_{\bar{F}} = \underline{9}.99$ (r,159) | | 4.70 | 5.31 | 5.68 |
| V-V | | .01 | .01 | .01 |
| M-V | | | .01 | .01 |
| M-NV | | | | n.s. |

CODE: V-V = Virginia - Verbal M-NV = Michigan - Nonverbal
M-V = Michigan - Verbal V-NV = Virginia - Nonverbal
$S_{\bar{F}}$ = Standard error for the mean of the groups representing levels of factor F (Form)
$\underline{9}$.95(r,159) = Tabled values of the Studentized Range Statistics for $(1 - \alpha) = .95$
$\underline{9}$.99(r,159) = Tabled values of the Studentized Range Statistics for $(1 - \alpha) = .99$
$S_{\bar{A}}$ $\underline{9}$.95(r,159) = Critical values for an ordered difference between two means, r steps apart
$S_{\bar{A}}$ $\underline{9}$.99(r,159) = Critical values for an ordered difference between two means, r steps apart

Figure 2

Verbal and nonverbal performance scores for Virginia and Michigan students by age level




The across age pattern of verbal performance in the two sample groups is nearly parallel. Also significant, although not related to Hypothesis No. 6, is the differential pattern of nonverbal performance scores across age. The consistent increase with age holds true of the Michigan students until age 13, but does not hold true for the Virginia students except on the overall age range. When the nonverbal scores are not separated by county, however, the effect is a consistent increase referred to on page 68.

Additional findings from the analysis of variance

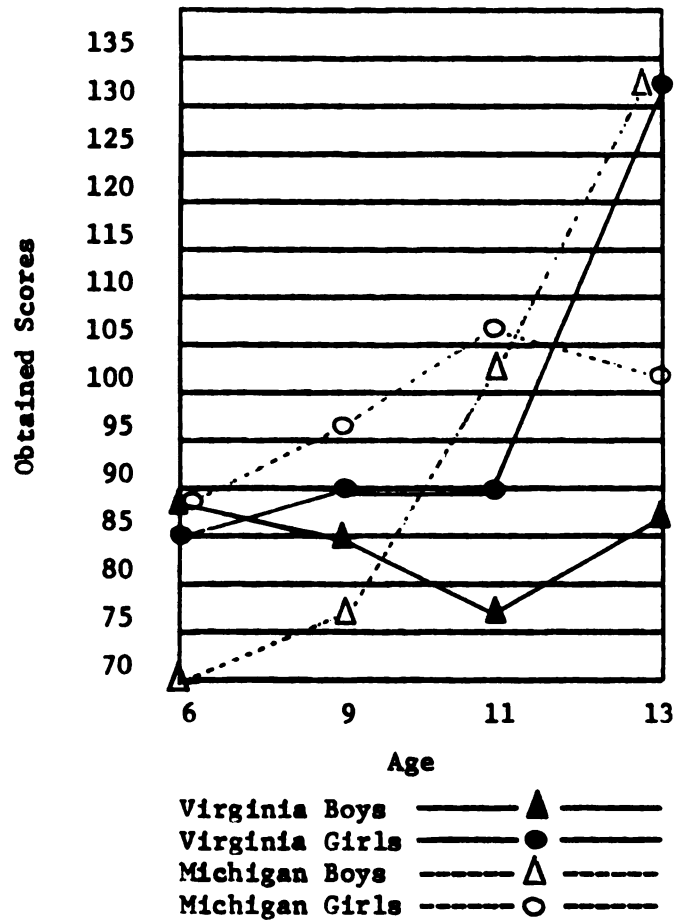
The preceeding references to the analysis of variance table and related techniques are those which dealt with the hypotheses formulated for this study. The remainder of this section elaborates upon the findings reported in the remainder of the ANOVA table (Table 4). This elaboration is sequentially given for each section of the table.

In the first section all of the results for main effects have previously been reported as well as the "County x Age" interaction. The interaction "Sex x Age" was insignificant indicating that when both population samples are considered together, girls tend to do better than boys at all age levels, and that increases in performance scores with age are similar in both sexes. Given the fact that the overall main effect of county was insignificant and that of Sex and Age were significant sources of difference, subsequent insignificance of the interaction of "County x Age" and "County x Sex" would seem to indicate that the obtained differences for both Sex and Age are consistent in the two populations studied. However, when the third order interaction of "County x Sex x Age" is analyzed the differences are highly significant. Thus, the inference would be that the interaction between Sex and Age differs significantly in the Prince Edward population in comparison to the Jackson population. Figure 3 is an attempt to graph this third order interaction.

This graph fails to clearly identify the major source of differences which account for the significance of the three way interaction. Thus, it was deemed desirable to employ some further means of analysis. The Neuman-Keuls technique was employed but the results, like the graph, were too diffuse to specifically identify the

Figure 3

Performance scores for Virginia and Michigan students, illustrated by sex and across age groups




factors involved. Finally, an attempt was made to innovate a technique. The result can best be described as a series of deviation tables. In essence, it is a procedure that allows the investigator to explain a complex interaction effect by setting up a series of expected values (based on results of the analysis of main effects and higher order interactions) and establishing the extent of the deviation of the actual obtained mean scores from the derived expected mean scores. This process is illustrated and the accompanying tables are presented in Appendix B.

The results of the analysis described in Appendix B indicate that the source of significance for the three-way interaction is the exceedingly low performance of thirteen year old Virginia boys and Michigan girls, and the relatively high performance of the six year old Virginia boys and nine year old Michigan girls.

The second subsection of Table 4 begins with the analysis of the main effect of "Scales". Using the usual "F" test procedure, this is significant at the .001 level. Under the adjusted degrees of freedom the "Scales" effect is significant at the .01 level. The four scales with their respective means in parentheses are: a) Fluency (32.8); b) Flexibility (21.7); c) Originality (18.3) and d) Elaboration (21.9). To further identify the source of the significant difference for performance on these scales, the Neuman-Keuls analysis is used with the results shown in Table 8.

These results would clearly indicate that the major source of the significant difference is the higher scores achieved on Fluency. Differences between the means of the other scales were not significant.

An examination of the scoring procedures from which these scale scores were derived would suggest that a certain degree of interscale dependency exists and is of the nature that Fluency would tend to be the highest score. The Flexibility score theoretically can equal the Fluency score but not exceed it. In actual practice, it would always tend to be lower. Originality and Elaboration have a similar dependency on Fluency in that the original and/or elaborated responses must first be recorded in the Fluency score. However, since each response can be scored in the excess of 1 for Originality or Elaboration the possibility of these scores exceeding the obtained Fluency scores exists. Thus,

Table 8

**Analysis of differences of pairs of means for performance scores on the four scales, using the Neuman-Keuls procedure**

| Scales | | $S_O$ | $S_{Flex}$ | $S_E$ | $S_{Flu}$ |
|---|---|---|---|---|---|
| Ordered Means | | 18.3 | 21.7 | 21.9 | 32.8 |
| Differences Between Pairs of Means | $S_O$ | | 3.4 | 3.6 | 14.5 |
| | $S_{Flex}$ | | | .2 | 11.1 |
| | $S_E$ | | | | 10.9 |
| $S_s$ = .29 | | | r=2 | r=3 | r=4 |
| $\underline{9}$.95 (r,477) | | | 2.80 | 3.36 | 3.69 |
| $\underline{9}$.99 (r,477) | | | 3.70 | 4.20 | 4.50 |
| $S_s \underline{9}$.95 (r,477) | | | 1.50 | 1.79 | 1.96 |
| $S_s \underline{9}$.99 (r,477) | | | 1.97 | 2.22 | 2.38 |
| $S_O$ | | | .01 | .01 | .01 |
| $S_{Flex}$ | | | | n.s. | .01 |
| $S_E$ | | | | | .01 |

CODE: O = Originality E = Elaboration
Flex = Flexibility Flu = Fluency
$S_s$ = Standard error for the mean of the group representing levels of factor s (Scale)
$\underline{9}$.95(r,477) = Tabled values of the Studentized Range Statistics for $(1 - \alpha) = .95$
$\underline{9}$.99(r,477) = Tabled values of the Studentized Range Statistics for $(1 - \alpha) = .99$
$S_s \underline{9}$.95(r,477) = Critical values for an ordered difference between two means, r steps apart
$S_s \underline{9}$.99(r,477) = Critical values for an ordered difference between two means, r steps apart

the nature of scoring techniques explains partially the obtained results but by no means adequately or completely explains the observed conditions. The lack of any significant differences between performance on the other three scales, Flexibility, Originality and Elaboration would seem to be independent of this factor. It would appear that the lack of differences reflects characteristics of the population samples.

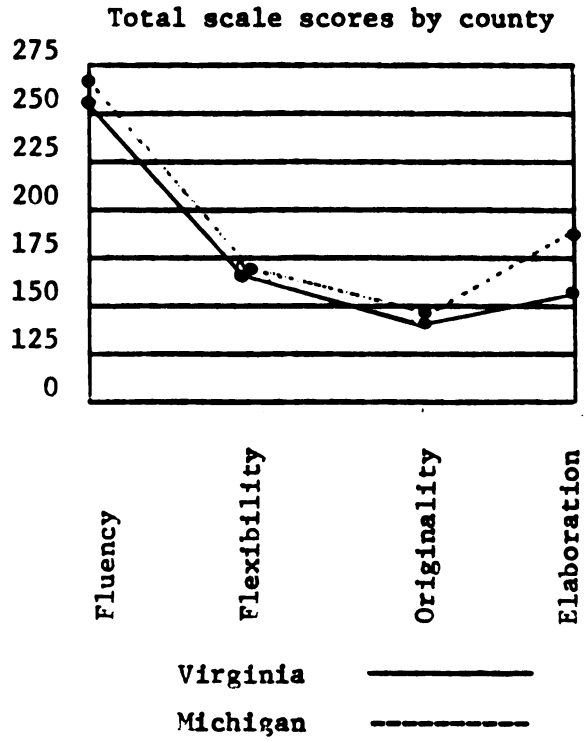
The two way interaction of "Scale x County" was insignificant, leading to the inference that the relative performance on these Scales, as reflected in the interpretation of the main effect of Scales, is consistent in both counties. However, because of its possible relevance to later interpretations, Figure 4 illustrates the Scale x County effect.

Although the "Scale x County" effect was insignificant, it does show the virtual equivalence of the two population samples on Fluency, Flexibility, and Originality scores, and the obvious superiority of the Michigan children on the Elaboration scale.

The results indicate that scale performance scores were relatively consistent in both sexes. This consistency is found to be less than for the county effects since the usual "F" test for the "Scale x Sex" interaction would indicate a .05 level of significance but the adjusted significance is .10.

The interaction effect of "Scales x Age" was significant at the .005 level after adjusting the degrees of freedom. This effect can be illustrated by the graphs shown in Figures 5 and 6. These graphs are sufficient to explain the source of the significance of the "Scale x Age" interaction. In part it is the function of an overall increase across all scales with age. However, the difference between achieved scale scores tends to be relatively consistent and in the same order as in the analysis of the main effect of Scales.

Figure 4




The only exceptions to this are the scores for the Elaboration scale which begin as the lowest at the six year old level, but ranks next to the highest for both 11 year olds and 13 year olds, with the difference proportionally greater for the latter age group. Thus, the primary contributor to the significance of the effect is the Elaboration scale, especially at the 6, 11 and 13 year old levels.

The three way interactions of "Scales x County x Sex," and "Scales x Sex x Age" were insignificant. Only "Scales x County x Age" interaction approachs significance with the regular "F" test showing significance at the .025 level; but after the appropriate adjustment of the degrees of freedom, this was reduced to .10. Again, because of its near significance and its possible help in interpreting the results reported for several of the hypotheses, this three way interaction, "Scale x County x Age," will be analyzed

Figure 5

Mean creativity scores by age and group

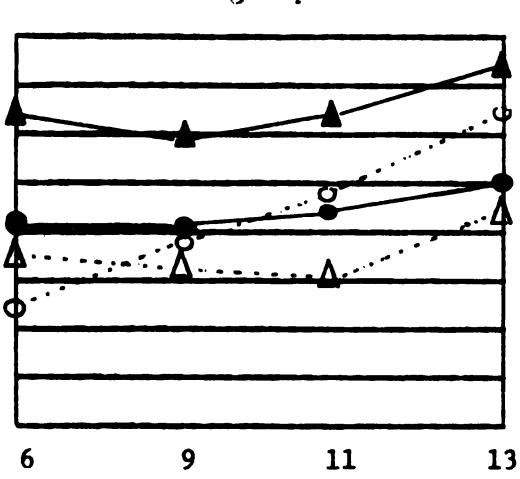
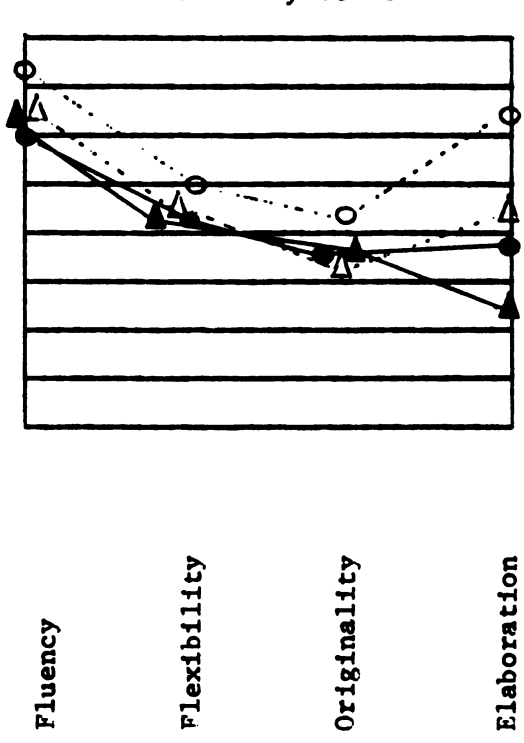



Figure 6

Mean age group scores on the four creativity scores




Fluency 6 ——▲—— Originality 11 ------ △ ------

Flexibility 9 ——●—— Elaboration 13 ------ ○ ------

in more detail. Figures 7, 8, 9 and 10 illustrate the differences in scale performance by the four age groups in each county. These four graphs illustrate for all scales at the four different age levels that the performance for the two population samples are virtually equal and parallel with the exception of the Elaboration scale for the three youngest age levels and the Flexibility scale at the fourth age level. The near significance of the "Scales x County x Age" interaction would seem to be a consequence of this shift of scale differential at the thirteen year old level.

In addition the fact that at the earlier age levels Virginia children demonstrated higher performance on most scales while at

Figure 7

Scale performance for 6 year old children

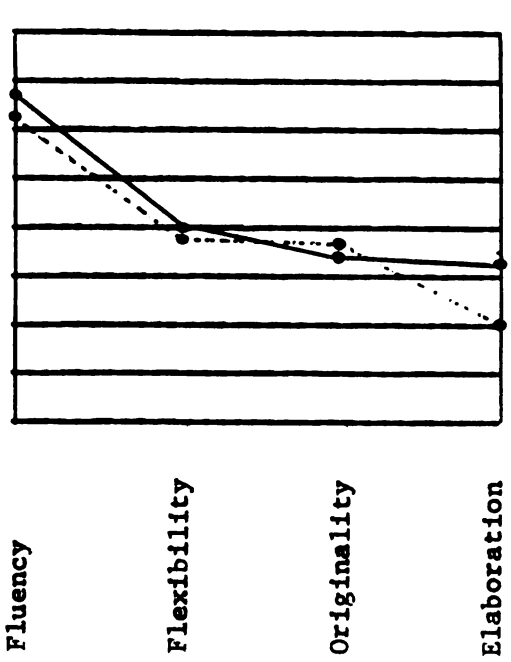



Figure 8

Scale performance for 9 year old children

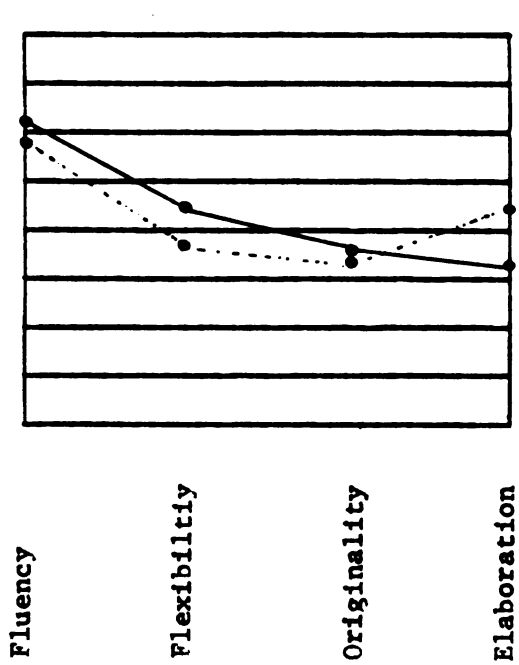



Figure 9

Scale performance for 11 year old children

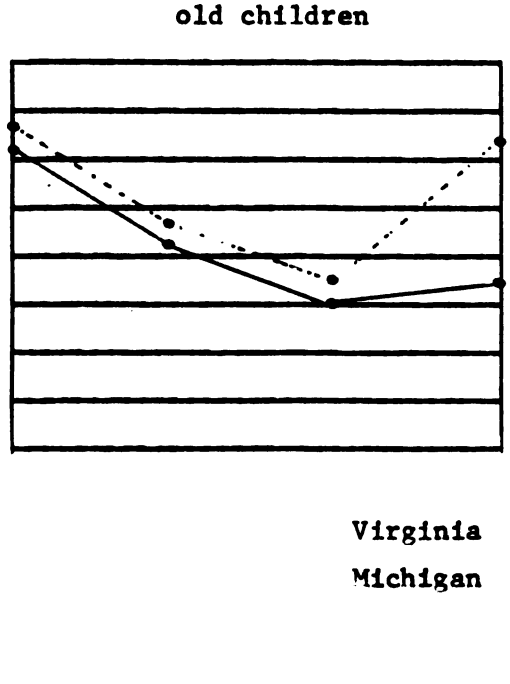
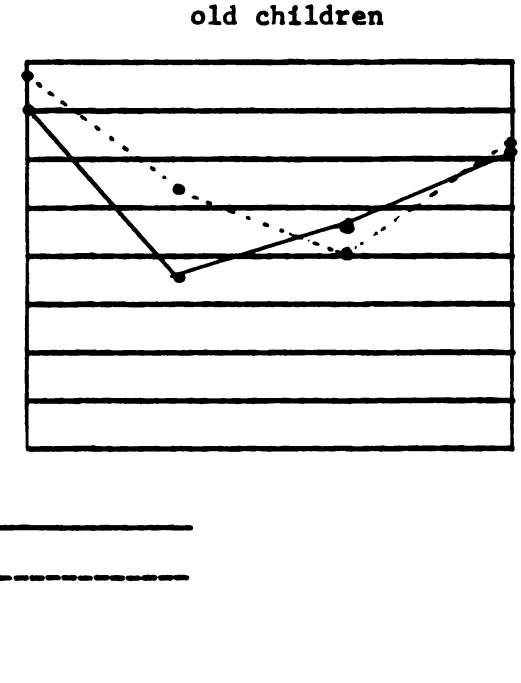



Figure 10

Scale performance for 13 year old children



Virginia ——————

Michigan - - - - - - - - - - - - -

the older age levels this superiority was reversed. This general trend had enough exceptions to also add to the magnitude of the significance of the effect.

The next subsection of the ANOVA table (Table 4, page 61) deals with the main effect of Form and related interactions. This main effect plus several of the interactions have been discussed earlier as they affected the disposition of specific hypotheses.

The interaction effect of "Form x Sex" was not significant, indicating that boys and girls reflected equal differences between verbal and nonverbal performance.

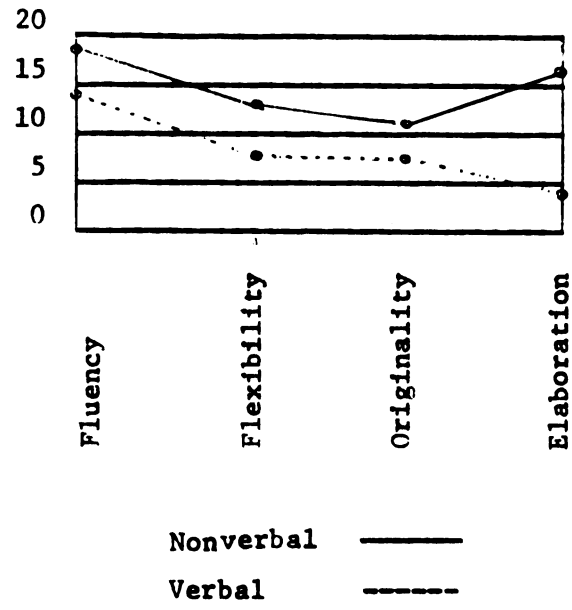
The three way interactions of "Form x County x Sex," "Form x County x Age," and "Form x Sex x Age," were not significant, nor was the four way interaction of "Form x County x Sex x Age."

The final part of the analysis of variance table begins with the "Scale x Form" interaction which was significant with the adjusted degrees of freedom at the .001 level. This interaction effect is graphed in Figure 11.

The primary source of the significance for the interaction effect of "Form x Scale" is the performance on the Elaboration scale. In the nonverbal form, Elaboration scores were only slightly lower, 2 points, than those for Fluency, and the scale in which the highest scores were achieved, and 4 points and 6 points higher than Flexibility and Originality, respectively. On the verbal scales, Fluency, Flexibility and Originality, scores are lower but parallel to the pattern on the nonverbal scale scores. However, the verbal Elaboration scores are the lowest of the four, ten points below the verbal Fluency score and four points below both verbal Flexibility and Originality.

Figure 11

Interaction effect of Scales x Form

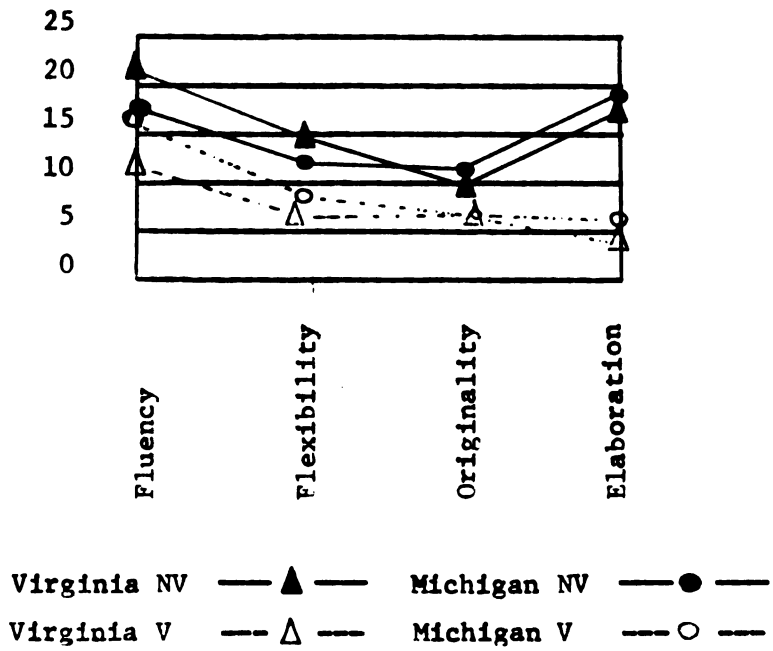



The three way interaction of "Scales x Form x County" also shows up as a source of significant difference. The adjusted "F" test results in a .05 level of significance. This interaction is graphed in Figure 12.

An interpretation of these results leads to a more specific identification of the sources of difference. From preceeding analyses of effects it was established that nonverbal scores were higher than verbal scores, and that this differential relationship was consistent across all scales with the exception of Elaboration. In the latter scale, the difference between verbal and nonverbal performance was maximized. Although the measured level of significance was .10, the interaction effect of "Form x County" showed that the Virginia students performed slightly better on the nonverbal tests than the Michigan students, and that the trend was reversed, and the differences slightly greater, on the verbal tests. Given these conditions, the source of significance

Figure 12

Interaction effect of Scales x Form x County




for the "Scales x Form x County" interaction seems to be the consequence of three factors: (a) the fact that the differences between the two counties on Fluency, both verbal and nonverbal, were relatively larger, Virginia greater than Michigan on the nonverbal, and Michigan greater than Virginia on the verbal scores; (b) the collapse of differences between counties on the Originality scale for both nonverbal and verbal (The scores for both populations were essentially the same for both nonverbal and verbal originality.); and (c) the higher performance scores achieved by the Michigan students on nonverbal Elaboration. None of these conditions, independently, is of the degree to have produced the significant difference but rather it is the combination of the three.

None of the remaining interaction effects prove to be significant. "Scale x Form x Sex" was insignifncant and "Scale x Form x Age" was significant until the adjusted degrees of freedom

reduced the significance level to .10. The computed significance level for the four way interactions of "Scales x Form x County x Age," "Scales x Form x County x Age," and "Scales x Form x Sex x Age," were all greater than .25, as was the five way interaction, "Scales x Form x County x Sex x Age." The results of the preceeding analysis and the subsequent disposition of the hypotheses will be presented in Chapter 6. In addition to these secondary analyses, Appendix C contains a descriptive analysis of scores achieved on the four creativity tasks which collectively made up the Abbreviated Form VII, Minnesota Tests of Creativity.

The primary function of the analysis reported in Appendix C was to determine if one or more tasks introduced biases that would differentially influence performance and confound the analysis reported in Chapter 4, especially with respect to such factors as sex or county. The conclusions derived from the analysis in Appendix C is that no such biases resulted. Other interpretations of the comparative task performance are included in the appendix.

# Chapter 5 - Secondary Analysis of the Data

## Introduction

This section deals with some treatments of the data that are not directly relevant to the testing of the hypotheses but can be of potential help in interpreting the results discussed in the last section as well as being interesting in their own right. There are three sets of supplemental data to be discussed: (a) a comparison of scores of the two samples analyzed in the previous section, Prince Edward County, Virginia and Jackson, Michigan, with performance scores for a sample from another county in Virginia; (b) a comparison of the I.Q. levels of Jackson, Michigan and Prince Edward County, Virginia, and (c) rank order correlations of I.Q. with the various total scores and form and scale scores and the comparison of these correlations with intertask and interscale correlations.

## Data from a neighboring Virginia county

It had originally been planned to run the complete battery of

tests at all four age levels on a matching sample from a county neighboring Prince Edward County in Virginia. In this way one could enhance control for the geographic or subcultural effect of Southern Rural Negro children and Northern Urban Negro children. Had this been accomplished, the Neighboring County sample and the Prince Edward County sample would have shared geographical or subcultural characteristics, while the Neighboring County and Jackson, Michigan, samples would have shared common school experiences, in terms of the involvement. Due to the temporary suspension of the Prince Edward County research project the resulting dates for the trip to Virginia to collect data included only the last half day of classes in the neighboring county. This meant that it was impossible to administer the total battery of Creativity tests to this sample. As a result only one of the four tasks, "Incomplete Figures," was administered to the three youngest age groups of the four age groups sampled from the other populations. Such limited data do not allow for making firm inferences concerning the subcultural factors. In addition, even though the tests were all administered by the principal investigator, the testing situations were not identical with those for the other samples and the effect of these differences are hard to determine in terms of increasing or decreasing performance scores. For these reasons these data are relegated to this secondary section rather than presented as an integral part of the analysis. The comparative scores are presented in Table 9 and illustrated in Figures 13, 14, 15 and 16. The limited data on the comparative scores for the three counties had several interesting patterns. Figure 13, depicting the mean scale scores on Figure Completion for the first grade level, shows a general equality among the mean scores for the children from the three counties. However, Figures 14 and 15 run counter to this pattern. The two Southern rural groups remain almost equal on all scales for both age levels while the Northern urban group deviates noticeably. The latter group, with a slightly lower mean score on Fluency and Flexibility at both age levels, achieves a limited superiority on

## Table 9

### Comparative scores for the three counties

| | | Fluency | | | Flexibility | | | Originality | | | Elaboration | | | Total Score | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | F | M | F&M | F | M | F&M | F | M | F&M | F | M | F&M | F | M | F&M |
| PE - 6 | $\bar{x}$ | 8.5 | 9.6 | 9.1 | 6.2 | 7.1 | 6.7 | 6.6 | 7.4 | 7.0 | 1.2 | 2.1 | 1.7 | 22.5 | 26.2 | 24.4 |
| | s | 2.46 | .7 | 1.85 | 2.3 | 1.5 | 1.95 | 3.1 | 1.6 | 2.43 | 1.23 | 2.38 | 1.90 | 5.97 | 3.36 | 5.08 |
| J - 6 | $\bar{x}$ | 8.1 | 8.6 | 8.3 | 6.5 | 7.6 | 7.0 | 8.9 | 7.2 | 8.1 | 4.5 | 2.4 | 3.5 | 28.0 | 25.8 | 27.0 |
| | s | 2.47 | 3.27 | 2.82 | 2.25 | 3.13 | 2.10 | 4.04 | 4.02 | 4.02 | 5.07 | 1.51 | 3.88 | 10.17 | 10.61 | 10.18 |
| NC - 6 | $\bar{x}$ | 9.82 | 6.7 | 8.4 | 7.64 | 4.4 | 6.2 | 8.6 | 5.8 | 7.3 | 2.81 | .7 | 1.9 | 28.8 | 17.6 | 23.8 |
| | s | .45 | 3.17 | 1.81 | 1.45 | 1.6 | 1.53 | 2.5 | 3.75 | 3.13 | 3.2 | 2.0 | 2.6 | 5.7 | 8.65 | 7.18 |
| PE - 9 | $\bar{x}$ | 8.6 | 8.9 | 8.7 | 7.0 | 7.6 | 7.3 | 5.56 | 4.9 | 5.3 | 4.33 | 2.9 | 3.7 | 25.5 | 24.6 | 25.1 |
| | s | 1.62 | 1.91 | 1.72 | 1.48 | 1.43 | 1.95 | 2.94 | 1.29 | 2.31 | 3.14 | 3.84 | 3.47 | 5.95 | 5.10 | 5.47 |
| J - 9 | $\bar{x}$ | 8.1 | 7.6 | 7.9 | 6.5 | 6.2 | 6.3 | 7.6 | 7.0 | 7.3 | 10.2 | 7.1 | 8.7 | 32.3 | 27.9 | 30.2 |
| | s | 2.64 | 2.80 | 2.67 | 1.98 | 2.75 | 2.33 | 3.50 | 2.72 | 3.10 | 8.58 | 4.09 | 6.85 | 12.25 | 6.85 | 10.07 |
| NC - 9 | $\bar{x}$ | 8.4 | 8.1 | 8.2 | 7.2 | 6.7 | 7.0 | 7.1 | 4.2 | 5.7 | 4.2 | 3.0 | 3.6 | 28.1 | 21.8 | 25.0 |
| | s | 1.85 | 2.43 | 2.14 | 1.62 | 2.05 | 1.84 | 2.84 | 1.1 | 1.97 | 4.15 | 3.3 | 3.73 | 4.45 | 6.81 | 5.63 |
| PE - 11 | $\bar{x}$ | 9.0 | 9.5 | 9.2 | 7.3 | 8.0 | 7.6 | 5.5 | 7.0 | 6.2 | 4.3 | 3.8 | 4.1 | 26.2 | 28.3 | 27.1 |
| | s | 1.95 | 1.27 | 1.66 | 1.67 | 1.15 | 1.46 | 2.84 | 2.26 | 2.65 | 3.85 | 2.70 | 3.31 | 7.55 | 4.42 | 6.28 |
| J - 11 | $\bar{x}$ | 7.9 | 8.8 | 8.4 | 6.6 | 6.8 | 6.7 | 6.8 | 7.1 | 6.9 | 12.3 | 10.2 | 11.3 | 33.6 | 32.8 | 33.2 |
| | s | 2.84 | 2.12 | 2.50 | 2.78 | 1.86 | 2.32 | 2.8 | 2.35 | 2.54 | 6.65 | 6.66 | 6.60 | 13.28 | 9.08 | 11.13 |
| NC - 11 | $\bar{x}$ | 8.2 | 8.9 | 8.5 | 4.8 | 7.4 | 7.0 | 4.5 | 5.7 | 5.1 | 6.8 | 6.0 | 6.4 | 26.2 | 28.0 | 27.0 |
| | s | 2.2 | 1.55 | 1.88 | 1.8 | 1.51 | 1.66 | 2.66 | 3.21 | 2.94 | 3.61 | 7.33 | 5.47 | 7.55 | 9.63 | 8.59 |

PE = Prince Edward County, Virginia

J = Jackson, Michigan

NC = "Neighboring County", Virginia

$\bar{x}$ = Mean

F = Female

M = Male

M & F = Female + Male Combined

s = Standard Deviation

Figure 13

Mean scale performance scores for first grade students on the figure completion task

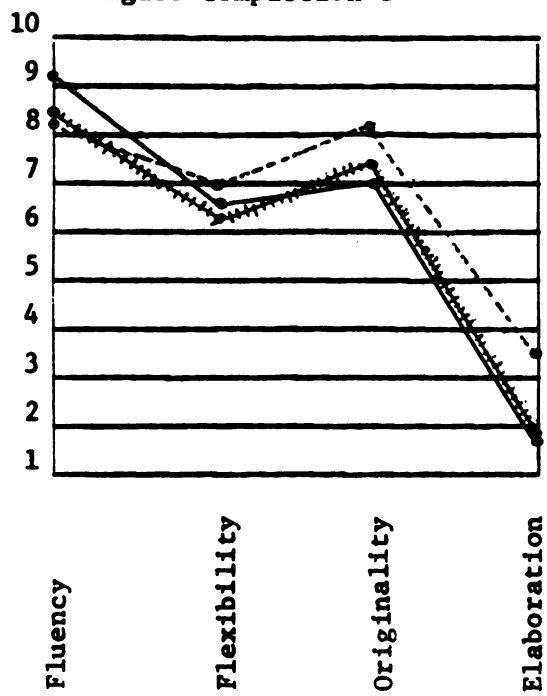



Figure 14

Mean scale performance scores for fourth grade students on the figure completion task

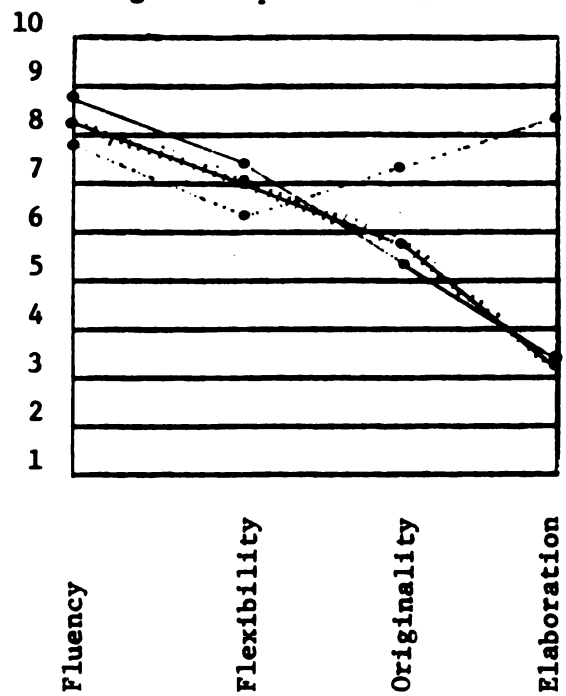



Figure 15

Mean scale performance scores for sixth grade students on the figure completion task

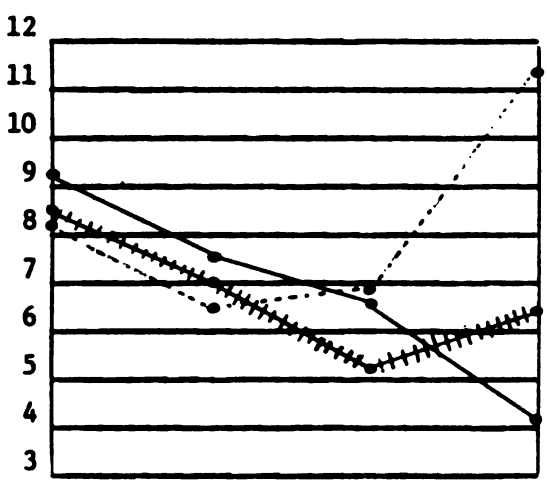
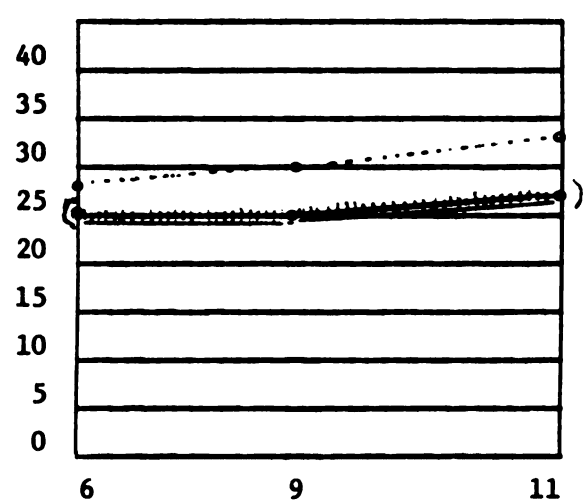



Figure 16

Mean total scores for age levels 6, 9 & 11 on the figure completion task




Prince Edward ————————

Jackson, Michigan - - - - - - - - -

"Neighboring County" ⁂⁂⁂⁂⁂⁂⁂⁂⁂

Originality and, with age, an increasingly sharp superiority in Elaboration. Given the fact that the neighboring county students had the typical continuous school experience shared by the Michigan sample the rationale for the differences of the older age level would seem to be weighted toward social-cultural differences. However, this reasoning requires the assumption of relatively identical school experiences in the two counties. This would be a risky assumption, for many reasons, none the least of which would be the limited creativity data and the irregular testing conditions for the neighboring county. Figure 16, plotting the mean total score for the Figure Completion task of the three sample groups at each of the three age levels, shows clearly the equality of the two Virginia groups with the Michigan children gradually increasing their superiority with age. From the immediately preceeding graphs, it is obvious that this latter effect is primarily a function of the differential performance on the Elaboration scale.

Certainly these sets of data are too limited and questionable to draw firm conclusions. At best it can be used to generate some questions that can only be resolved by further research concerning the comparative effects of social cultural variables on performance on creativity tasks and more specifically a differential effect of scale scores. At the same time the data reflect an adequately consistent pattern of similarity between the two Virginia samples strongly suggesting that the major factor producing the differential effect observed in this study may be of a more general social-cultural origin than the more specific factors of differences in formal educational experience.

## Creativity and intelligence

One of the major issues that has emerged from the recent

literature on research in creativity has been the nature of its relationship to intelligence as measured by the standardized tests and reported in the I.Q. score (Getzels and Jackson, 1962; Torrance, 1962, 1963b; Ripple and May, 1962). The various reports have labeled the relationship as (a) minimal, insignificant or even negative, (b) low, but significant up to a certain threshold beyond which the two variables are independent and (c) significant in a normal population implying that reports of independence are a statistical function of using attenuated samples with little variation in I.Q.

This study can add little clarity to this confusion. In the first place a relatively attenuated sample is again involved. Instead of a sample characterized by high I.Q., as has been the case in other studies, this study has samples characterized by generally below average I.Q.'s. In the second place, the I.Q. data available are based on various intelligence tests given at different age levels. To have grouped these together and made inferences from resulting correlations with creativity scores would have been a questionable procedure. Thus, the data described below are for the purpose of facilitating the interpretation of the results described in the previous chapter, and exploring possible facets of the relationship between the variables of "creativity" and "intelligence" for tentative insights to generate further investigation.

Table 10 presents mean I.Q. scores for Michigan and Virginia students by age levels. The sources of I.Q. are mixed to such an extent, even within age groups, that no attempt will be made to specify the source. It should, however, be noted that the I.Q. scores for the 13 year old Jackson students are derived from the Verbal Reasoning Scores on the Differential Aptitude Test. It is interesting to note that the six year old children in both Virginia and Michigan were essentially equal in performance on I.Q. tests. Following this, the reported average I.Q. of the Michigan students

Table 10

Comparative I.Q. scores for Michigan and Virginia students by age levels

| | 6 | 9 | 11 | 13 |
|---|---|---|---|---|
| Prince Edward | 88 | 83.1 | 73.7 | 87 |
| Jackson, Michigan | 87.6 | 97.2 | 94.6 | 91 |

rises as age and school experience increase; whereas in Virginia an increase in age, without a concommitant increase in school experience, produces a declining average I.Q. for the 9 and 11 year old age groups. However at the 13 year old level, the I.Q. scores are nearly equal. It should be kept in mind that the 13 year old children in Prince Edward County had attended grades K-3 before the public schools were closed.

The next set of data to be presented is that resulting from rank order correlations between I.Q. scores and the scores on the "Minnesota Tests of Creativity, Abbreviated Form VII." These correlations were computed for the total scores on the four scales and for the two forms of the test. Wallach and Kogan (1965) have pointed out a flaw that is contained in much of the analysis on the relationship between I.Q. and Creativity:

> The 'creativity index' used has varied for different studies, consisting in each case of a composite score based on the results of two or more different procedures. In the majority of these studies, the intelligence measure utilized has been an I.Q. score, a variable from which age differences have been partialled out. Age differences have not, in turn, been partialled out from the creativity scores. The effect of relating the creativity and intelligence measure under these conditions, if there is a reasonably extensive age range, is to underestimate the degree of relationship between them. This is the case since age has been partialled out of one variable but not the other. The evidence in question, then, cannot be said

to convince one of the relative independence of 'creativity' and 'intelligence'. (pages 8-9)

To avoid this flaw the rank order correlations were computed separately for each age group separated by "county". This then necessitates the questionable practice or summing the rank order correlations from several groups and computing the arithmetic mean correlation for the total correlations where age or county groups were combined. Such a procedure would not be acceptable for drawing conclusions but is satisfactory for this exploratory effort. As has been discussed before, it was necessary to use the rank order technique because of the diversity in type and time of I.Q. measurement. This same factor argued against the combining of counties at the same age level. It is also true that the degree of correlation obtained is seriously attenuated by the low reliabilities of the Creativity scores.

The results of the analysis of the relationship between I.Q. and Creativity are presented in Table 11. The most significant feature of this set of data is the higher correlations between Creativity scores and I.Q. scores for the Prince Edward sample. It is not implied that all these differences in obtained correlations are significant differences; rather that the "significance" lies in the pervasiveness or consistency of the trend.

This is the only consistent pattern to emerge from the above correlations. The correlation between I.Q. and Creativity Total Score does not systematically increase or decrease with Age, and the same fluctuation is observed when the correlations are computed separately by county. When the correlations are computed by total scale score, Flexibility and Elaboration are significantly correlated with I.Q. The latter was the only scale to show a significant correlation with I.Q. for the Jackson sample, while only Originality failed to show a significant correlation for the Prince Edward County group. When these are inspected by age, the eleven year old children

## Table 11

Rank order correlations between I.Q. and certain performance scores on the Minnesota tests of creativity

| | Total Population | Virginia | | | | Michigan | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | | 6 | 9 | 11 | 13 | | 6 | 9 | 11 | 13 | | 6 | 9 | 11 | 13 |
| Total Score | ***** .21 | ** .31 | .26 | .01 | .29 | ***** .36 | ** .48 | .30 | .14 | ** .51 | .09 | .15 | .24 | .09 | -.05 |
| Nonverbal Sub-Total | **** .20 | * .28 | *** .37 | .06 | .20 | **** .32 | .34 | .43 | .09 | .43 | .10 | .24 | .34 | -.13 | .05 |
| Verbal Sub-Total | *** .17 | .16 | .04 | .06 | .42 | *** .27 | .23 | -.05 | .24 | .54 | .09 | .09 | .09 | .09 | -.08 |
| Fluency | .09 | .14 | .09 | -.06 | .21 | ** .20 | .24 | .07 | -.14 | *** .45 | .00 | .10 | .10 | -.12 | -.05 |
| Flexibility | *** .22 | .22 | .23 | .19 | .23 | *** .26 | .19 | .27 | .26 | ** .42 | .18 | .24 | .20 | .07 | .04 |
| Originality | .01 | .05 | .02 | -.21 | .19 | .07 | .07 | .02 | -.11 | *** .49 | -.05 | .04 | .02 | -.13 | -.12 |
| Elaboration | ****** .36 | .47 | .49 | .06 | .46 | ****** .50 | **** .67 | **** .63 | -.14 | *** .48 | *** .24 | .29 | ** .40 | .14 | *** .44 |
| N's = | 149 | 31 | 37 | 40 | 41 | 68 | 15 | 14 | 18 | 21 | 81 | 16 | 23 | 22 | 20 |

* Significant at .10 level
** Significant at .05 level
*** Significant at .025 level
**** Significant at .01 level
***** Significant at .005 level
****** Significant at .0005 level

again show a considerably lower correlation between I.Q. on three of the scale scores than do the other three age levels. Flexibility is the exception. This latter scale showed the greatest consistency in its correlation with I.Q. across all age levels. When the correlations were compared relating Form with I.Q., the differential effect by county was of roughly the same order for both verbal and nonverbal. When the dichotomy by form is further broken down by age this consistency of differences between the counties breaks down.

One final comparison can be made with the available data. One of the criticisms made of the work done in defining and assessing creativity is that it is depicted as a single, unified dimension that can be differentiated from "intelligence" as measured by the standardized I.Q. test. This criticism has been made most explicitly by Wallach and Kogan (1965):

> Perhaps the most widely publicized recent research in this area is the volume Creativity and Intelligence by Getzels and Jackson (1962). These authors worked with five procedures labeled "creativity tests" and considered the performance of a large sample of students (class range, sixth grade through senior year of high school) on these and on one or another I.Q. measure. With 292 boys and 241 girls under consideration, all five of the creativity tests were significantly ($p < .05$) correlated with I.Q. for the boys, and four out of the five creativity tests were significantly ($p < .05$) correlated with I.Q. in the case of the girls. Two of those five correlations for the boys were not only significant but substantial (.37 and .37); analagously, for the girls, two of those five correlations were substantial as well as significant (.39 and .37). As a first point, then, one can question the degree to which the tests of creativity are independent of general intelligence.
>
> Consider next the relationship among the creativity tests themselves. Inspection reveals that the five tests in question virtually are no more strongly correlated with one another than they are with intelligence, and that this is true for both sexes. Thus the average correlation for boys is .26 between the creativity tests and I.Q., and is .28 among the

> creativity tests themselves. Thus, if one wishes to argue from these data that a modicum of commonality underlies the creativity tests, then one must admit that almost the same degree of commonality also extends to the intelligence measure, so that no evidence is found for conceiving of a psychological dimension of creativity as existing apart from general intelligence. If, on the other hand, one wishes to propose from these data that the creativity indices possess much variance that is distinct from differences in general intelligence, then one must admit that these creativity tests are in just about the same degree independent of one another, so that no evidence exists for conceiving of a single, unified dimension that would be appropriately labeled creativity after the manner of the concept of general intelligence or "G". (p. 3)

It should be noted that Wallach and Kogan's questioning of the legitimacy of a single operational construct of creativity, the product of summing the performance scores of five creativity tests, has in part been recognized by the design of this study. Although the analysis of mean effects utilized a similar "global" score, i.e. the total score achieved on each of four tests, added together for one "grand" total score, it is also true that the effects were also analyzed using total scale scores and scale scores by verbal or nonverbal form. By utilizing analagous analysis procedures on the data obtained from this study it is not only possible to see if Wallach and Kogan's results are replicated, but to also see if treating the performance scores achieved on different tasks as additive by scale is legitimate, and to further explore the resulting relationships when form, verbal or nonverbal, is included.

Wallach and Kogan state that results similar to those achieved by their analysis of the Getzels and Jackson data are found for other such studies.

Turning to the work of Torrance and his associates, the developers of the tests utilized in this study, Wallach and Kogan (1965) report:

> No data have been presented by the Torrance group that would permit one to evaluate whether and to what extent the various creativity procedures they utilize intercorrelate more strongly with one another than the degree of their separate correlations with general intelligence. However, the creativity instruments used by the Torrance group seem to possess about the same degree of diversity as those employed in the Getzels-Jackson, Cline-Richards-Needham, Cline-Richards-Abe, and Flescher studies, where the creativity procedures correlated to about the same degree with general intelligence as they correlated among themselves, or actually correlated more strongly with general intelligence than they correlated among themselves. From the direction of this evidence, therefore, we have to wonder whether the correlational situation would not be rather similar in the case of the Torrance group's materials. (p. 9)

The following set of data, given the various limitations of the study, should illuminate this issue. It should be noted that the n's are different for the intercorrelations among creativity indices than for those of the correlations between the I.Q. and creativity indices. The former n's represent the total sample used for the analysis of variance purposes. The latter n's are the major proportion of the total sample for which I.Q. data was available. If one assumes that those eliminated for want of I.Q. data constitute a randomly selected group, and no evidence exists to indicate otherwise, then this factor should not seriously contaminate the results.

A replication of the analysis applied by Wallach and Kogan to the Getzels and Jackson data, is perhaps the best place to begin. As was reported earlier in this chapter, the correlation between total score and I.Q. for the total population was .21 which was significant at the .05 level. The following table gives the intercorrelations between the four different tasks of the creativity battery. The confidence intervals for the correlations are given in parentheses.

Table 12

Inter-task correlations total sample
N = 171
(confidence intervals)

| | Circles | Product Improvement | Uses |
|---|---|---|---|
| Figure Completion | .40 (.27-.52) | .27 (.13-.40) | .26 (.11-.39) |
| Circles | | .00 (-.15-.15) | .09 (-.06-.24) |
| Product Improvement | | | .37 (.23-.49) |

The correlations for each sub-test with the Total Score were as follows:

| | | |
|---|---|---|
| Figure Completion | .70 | (.62-.85) |
| Circles | .68 | (.59-.75) |
| Product Improvement | .55 | (.43-.64) |
| Uses | .64 | (.55-.72) |

Table 12 indicates a limited acceptance of Wallach and Kogan's assumptions concerning the creativity tests of Torrance and his associates, or at least the four tests utilized in this study. Only two of the intercorrelations for the task show confidence intervals which exceed the correlation of I.Q. with total Creativity Score. Two of the intercorrelations are of generally equal magnitude to the I.Q.-Creativity relationship and the remaining two are less. It is also interesting to note the correlations of the four individual creativity tasks with I.Q. They are as follows:

| | | |
|---|---|---|
| Figure Completion | .14 | (not significant) |
| Circles | .23 | (significant at the .05 level) |
| Product Improvement | .14 | (not significant) |
| Uses | .11 | (not significant) |

It should be noted that those correlations for which confidence intervals are used are product-moment correlations while significance levels are used for rank-order correlations.

Because there was a substantial difference between the I.Q.-Creativity correlation for the two counties, .36 for Virginia and .09 for Michigan, the above correlations should also be reported separately for the two samples.

Table 13

Inter-task correlations for Virginia students
N = 85
(confidence intervals)

| | Circles | Product Improvement | Uses |
|---|---|---|---|
| Figure Completion | .48 (.30-.63) | .21 (.01-.40) | .33 (.13-.51) |
| Circles | | .19 (-.03-.39) | .32 (.11-.50) |
| Product Improvement | | | .53 (.35-.66) |

Table 14

Inter-task correlations for Michigan students
N = 90
(confidence intervals)

| | Circles | Product Improvement | Uses |
|---|---|---|---|
| Figure Completion | .46 (.27-.61) | .25 (.04-.43) | .15 (-.06-.35) |
| Circles | | -.04 (-.25-.17) | .11 (-.10-.31) |
| Product Improvement | | | .24 (.03-.43) |

Tables 13 and 14 show a more consistent agreement with the Wallach and Kogan argument. In Table 13 none of the confidence intervals for the inter-task correlations exceed the .36 correlation between I.Q. and creativity for the Virginia population. In Table 14 only the intercorrelation between Figure Completion and Circles shows a confidence interval that exceeds the .09 correlation between I.Q. and Creativity for the Michigan children. The correlation of each task with total score is presented in Table 15 for the

two sample subgroups.

Table 15

Correlations of individual creativity task scores with total creativity scores for Virginia and Michigan children

| Task | Virginia | Michigan |
|---|---|---|
| Figure Completion | .69 (.56-.79) | .70 (.57-.79) |
| Circles | .81 (.72-.87) | .69 (.56-.78) |
| Product Improvement | .60 (.44-.72) | .49 (.31-.64) |
| Uses | .71 (.58-.80) | .61 (.47-.73) |

Table 16 shows the correlations of I.Q. with each of the four task scores for the two counties.

Table 16

Correlation of I.Q. with each of the four Creativity Task Scores

| | Virginia | Michigan |
|---|---|---|
| Figure Completion | .01 | .26 |
| Circles | .42 | .07 |
| Product Improvement | .19 | .09 |
| Uses | .21 | .03 |

Table 15 shows that the correlation of the individual tasks with total score do not vary to any great extent between the two counties, but Table 16 indicates that there is a differential effect by county on correlations between I.Q. and task scores. Circles, Uses, and Product Improvement scores showed a relatively substantial correlation with I.Q. for Prince Edward County children, and a near zero correlation for the Michigan children. Figure Completion scores reversed this pattern with correlations with I.Q. near zero for the Virginia children and fairly substantial correlations for the Michigan children.

These correlations can also be viewed in terms of dichotomizing the total score into nonverbal and verbal scores, and comparing the correlations of these two subtotals and I.Q. with the respective correlation between the two nonverbal and two verbal tasks. For the total sample the correlation between I.Q. and nonverbal Creativity is .20 while the correlation between Figure Completion and Circles is .40. Verbal Creativity and I.Q. were correlated at .17 for the total population as compared to the .37 correlation between Product Improvement and Uses. If these are further separated into figures for Virginia and for Michigan the results are:

| | | | |
|---|---|---|---|
| (a) | Virginia | I.Q. x Nonverbal | .32 |
| | | Figure Completion x Circles | .48 |
| | | I.Q. x Verbal | .27 |
| | | Product Improvement x Uses | .53 |
| (b) | Michigan | I.Q. x Nonverbal | .10 |
| | | Figure Completion x Circles | .46 |
| | | I.Q. x Verbal | .09 |
| | | Product Improvement x Uses | .24 |

In all cases the relevant intertask correlations are higher than the appropriate form creativity scores with I.Q.'s. The magnitude of this difference can be considered substantial for all but one, the nonverbal scores of the Virginia children.

One of the major implications in Chapter 3 concerning the analysis of creativity scores was the desirability of utilizing scale scores (Fluency, Flexibility, Originality and Elaboration) as opposed to scores for the individual tasks or even total scores. This preference is reflected in the analysis of variance design utilized for the major analysis. Thus it seems wise to submit the data, organized within scales, to the analysis suggested by Wallach and Kogan (1965).

The first method of describing the relationship between "creativity" scale scores utilizes the total scale scores summed across the four tasks. These relationships are depicted in Table 17.

Table 17

Intercorrelations between total scale scores
( ) = confidence intervals
Correlation of I.Q. with creativity total score = .21

| | Flexibility | Originality | Elaboration |
|---|---|---|---|
| Fluency | .78 (.72-.84) | .64 (.54-.72) | .41 (.28-.53) |
| Originality | | .66 (.57-.74) | .48 (.36-.59) |
| Elaboration | | | .42 (.29-.54) |

It is also interesting to note two other sets of relationships based on this grouping of data. The first of these is the correlations of the four scale scores with the total score. These are as follows:

| | | |
|---|---|---|
| Fluency | .83 | (.77-.87) |
| Flexibility | .86 | (.77-.88) |
| Originality | .80 | (.74-.85) |
| Elaboration | .79 | (.72-.83) |

The other relationship is that between the scale scores and I.Q. These as previously reported on page 92 were:

| | | |
|---|---|---|
| Fluency | .09 | (not significant) |
| Flexibility | .22 | (significant at .005 level) |
| Originality | .01 | (not significant) |
| Elaboration | .36 | (significant at the .0005 level) |

As with the data grouped by task scores, this description should be presented separately for the two "counties" due to the difference between the correlations of total score with creativity for the two groups.

Table 18

Intercorrelations between Total Scale Scores for Virginia children
( ) = confidence intervals

| | Flexibility | Originality | Elaboration |
|---|---|---|---|
| Fluency | .76 (.65 - .84) | .68 (.55-.78) | .40 (.20-.56) |
| Flexibility | | .73 (.61-.81) | .54 (.37-.68) |
| Originality | | | .51 (.33-.65) |

Table 19

Intercorrelations between Total Scale Scores for Michigan children
( ) = confidence intervals

| | Flexibility | Originality | Elaboration |
|---|---|---|---|
| Fluency | .81 (.72-.87) | .67 (.54-.77) | .43 (.24-.59) |
| Flexibility | | .70 (.58-.79) | .45 (.27-.60) |
| Originality | | | .33 (.13-.51) |

Tables 20 and 21 show the differential correlations for the two sample groups between the individual total Scale Scores and the Total Creativity Score and between I.Q. and the Total Scale Scores.

Table 20

Correlations for Michigan and Virginia children between total scale scores and total creativity scores
( ) = confidence intervals

| Scale | Virginia | Michigan |
|---|---|---|
| Fluency | .82 (.73-.88) | .85 (.78-.90) |
| Flexibility | .87 (.81-.91) | .85 (.78-.90) |
| Originality | .84 (.77-.89) | .78 (.68-.85) |
| Elaboration | .80 (.71-.87) | .77 (.67-.84) |

Table 21

Correlations for Michigan and Virginia children between scale scores and I.Q.
( ) = significance levels

| Scale | Virginia | Michigan |
|---|---|---|
| Fluency | .20 (.05) | .00 (ns) |
| Flexibility | .26 (.025) | .18 (ns) |
| Originality | .07 (ns) | .05 (ns) |
| Elaboration | .50 (.0005) | .24 (.025) |

This last group of data, dealing with Creativity Scale Scores, also raises some interesting questions and adds a few dimensions to previously raised questions. The major reason that this portion of the analysis was undertaken was to apply the criticisms of Wallach and Kogan (1965) to data derived in this study. The focal issue dealt with the disparaging of the "creativity" construct in the global sense derived from tasks such as those employed by Guilford, Getzels and Jackson, and Torrance. The purpose was not to refute

the Wallach and Kogan assertions but rather to investigate this phenomena when scale scores are utilized as the basic referent as opposed to task scores. The previous analysis of task scores for the total sample did not clearly refute or substantiate the assertions under consideration. Because of the substantial differences between the Virginia and Michigan samples with respect to the degree of relationship between I.Q. and Creativity, the analysis was then computed separately for these two groups. In an important sense this provided support for the Wallach and Kogan claims.

When interscale correlations are compared with the I.Q. x Total Score correlations the intercorrelations among the scales are significantly higher than the I.Q. x Total Score correlation. At no point do the confidence intervals for these interscale correlations include a .21 correlation between I.Q. and total score of creativity. It is also interesting to note the Elaboration scale which shows the lowest correlation with the other scales, also demonstrates the highest correlation with I.Q.(pp. 100-101). These results raise some questions as to the validity of the Elaboration scale in the creativity battery. Such considerations, however, are beyond the scope and concern of this study.

As Tables 18 and 19 ( p. 101) show, the Virginia interscale correlations, while generally higher than the correlation of I.Q. and Total Creativity score, are not always significantly higher as determined by the range of the confidence interval. Conversely, the Michigan sample which has a more normal I.Q. distribution shows all six interscale correlations significantly higher than the correlation of I.Q. and Total Scale Creativity. Table 21 shows that the ordering of scale scores with respect to their correlation with I.Q. would produce the same order in both counties. However, the correlations do not run parallel to each other; Fluency and Elaboration show a decidedly higher correlation with I.Q. for the Virginia students than the Michigan students.

The correlations between the two forms (Verbal and Nonverbal) of the Scale scores were as follows:

| | |
|---|---|
| Fluency | .00 |
| Flexibility | .14 |
| Originality | .17 |
| Elaboration | .17 |

Given these low correlations between verbal and nonverbal scale scores, one would infer that had interscale correlations been computed separately by form, they would have been higher than those reported in Tables 17-19.

These low correlations between forms of the four scales would also suggest that the grouping of form scores into a single score is a questionable procedure. Further work with larger and more typical population samples will be required to ascertain the specific grouping of scores that is most valid for analytical purposes.

## Chapter 6 - Discussion

The final chapter of this study is primarily concerned with interpreting the results of the analyses reported in Chapter 4 and deriving appropriate inferences as to their meaning or implications.

### Review

Before discussing the various conclusions and implications derived from this study, it is best to review a summary of the disposition of the research hypotheses and characteristics or aspects of the study which impose various limitations on the conclusions drawn.

Summary of primary statistical analysis:

1. The performance scores of the Prince Edward County, Virginia, children will be significantly higher than those scores of the Jackson, Michigan children.
   Disposition: rejected.
   (a) There will be no significant differences in the performance of the Virginia and the Michigan students in the six year old group.

Disposition: rejected. The Virginia students achieved significantly higher performance scores, significant at the .05 level.

(b) The fourth grade (nine year old) Virginia students will do significantly better on the tests of Creativity than will the fourth grade Michigan students.
Disposition: rejected. The slight difference was in the right direction but was not statistically significant.

(c) The sixth grade (11 year old) Virginia students will do significantly better on the tests of creativity than will the sixth grade Michigan students.
Disposition: rejected. The Michigan students achieved higher scores than did the Virginia students and the difference was statistically significant at the .01 level.

(d) The eighth grade (13 year old) Michigan students will do significantly better on the tests of Creativity than will the eighth grade Virginia students.
Disposition: rejected. The observed difference was in the right direction, but was not statistically significant.

2. The factor of age will cause significant differences in performance scores. (An increase in age will be accompanied by an increase in performance scores.)
Disposition: accepted. (Significant at the .001 level.) It should be noted however that there was not significant differences between all age intervals.

3. There will be no difference in performance when compared by sex.
Disposition: rejected. Girls scored significantly higher than boys. (Significant at the .05 level.)

4. The mean scores of the total study population will be significantly higher on the nonverbal forms in comparison to the verbal form.
Disposition: accepted. (Significant at the .005 level.)

5. The difference between verbal and nonverbal performance will decrease significantly with an increase in age.
   Disposition: rejected. Actually the differences increase with age to such a degree that "Form x Age" was significant at the .005 level.
6. (a) Virginia students performance scores on the nonverbal tests of Creativity will be significantly higher than the scores attained by the Michigan students.
   Disposition: rejected. Difference was in the right direction, but not statistically significant.
   (b) The Michigan students performance scores on the verbal tests of Creativity will be significantly higher than the scores attained by the Virginia students.
   Disposition: accepted. (Significant at the .01 level.)

As has been mentioned frequently in the preceeding chapters, the assessment of creative behavior by relatively objective paper and pencil techniques has been frequently criticized. These criticisms have focused on both the questions of validity and reliability. The issue of validity has been discussed previously within the context of theoretical constructs and operational definitions of creativity. Reliability figures obtained by Torrance and his associates and other users of the instruments have been reported. In addition, the results of a reliability study connected with this study have been given. If one was to draw a general conclusion from both sources of reliability data, it would be that the Minnesota Tests of Creative Thinking, Abbreviated Form VII (Torrance 1963), demonstrate a marginal reliability, high enough to allow limited and tentative inferences to be drawn for research purposes, but not adequate to permit their use as a means of evaluation, either of programs or individuals. With a few possible exceptions, the results shown in Table 22, page 135, Appendix A, show no consistent patterns or trends that lead to conclusions other than those previously reported from other literature.

The problem of low reliability of the criterion measures is but one of several limitations of this study that restricts the validity and

generalizability of the conclusions drawn in this chapter.

Another limitation is that of inability to control for examiner effect. Actually, there are two dimensions to this condition or restriction. The first of these dimensions exists as a consequence of the relatively unstructured or ambiguous nature of the testing stimuli. Generally speaking, it is agreed that the more ambiguous an assessment task, the greater the possibility for those variables associated with testing conditions and the test administrator to affect the responses elicited. The types of tasks utilized for this study would have to be considered as well along the continuum toward the unstructured pole. Added to this is the condition that these tasks were administered to the subjects by a total of six different individuals. This fact has been elaborated in Chapter 3. Although all of the examiners were given training sessions on the administration of the various tasks and all administered several practice tests under the supervision of the principal investigator, it is only prudent to assume the presence of examiner effect. Compounded with this factor are the possible effects of male-female and Caucasian-Negro examiner dichotomies. The principal investigator was the only male of the six examiners. The two individuals who assisted the investigator in Prince Edward County, Virginia, were Negros. The three who administered tests in Jackson, Michigan were Caucasians. It should also be remembered that the entire sample was comprised of Negro students. Unfortunately, the procedures for collecting the data, procedures which were dictated by factors beyond the control of the investigator, did not allow for the design of any analysis that approaches a systematic examination, or even control, of the various examiner effects. An attempt to assess the effect of individual differences in test administration is negated for the group-administered tests because one examiner gave group tests in Virginia and the other in Michigan. Therefore these differences are compounded in the between-county effects. The between-county effects also obscure the Negro-Caucasian examiner effect on the individual tests, as the Virginia students were examined by Negro individuals and the Michigan students by Caucasian examiners.

It should be noted that such an arrangement did have the advantage of paralleling the racial character of the pupil-teacher relationship the children had experienced previously and were experiencing at the time of the study. The fact that in both Virginia and Michigan one examiner tested primarily six year old students and the other individual primarily tested the nine and eleven year old students results in these examiner differences being embedded in the between age differences found in the sample.

This failure, or inability, to control for the effects of examiner, along with the factor of restricted reliability, are the two major weaknesses of this study. A major weakness is also the study's failure to control for social-cultural differences between the populations sampled. As has been discussed previously the original design which called for a matching sample from a county neighboring Prince Edward County would have provided adequate controls. The inability of the investigator to carry out this phase of the design does not allow one to distinguish social-cultural factors, primarily Southern rural-Northern urban differences from differences in school experiences, the independent variable of the study, in explaining the observed differences in performance on creativity tests. The limited data that was obtained from the neighboring county suggests that the social-cultural differences do effect performance on the criterion tasks. Therefore it is necessary to caution the reader to view any conclusions or inferences reported in this chapter as being highly tenuous. The major value of this study would seem to lie in its identification of several areas of concern that merit future rigorous research. Keeping the above limitations in mind, the next step in the systematic examination of the results of the analysis of variance design and the interpretation of the disposition of the formulated hypotheses.

## Discussion of hypotheses

Hypothesis No. 1, as tested by the main effect of County, is

probably the most pervasively crucial hypothesis of the study. The basic purpose of the study was to analyze the effects of lack of formal schooling upon performance on tasks of creative thinking. This was done from within the context of a rationale that depicted the process of formal education, as generally provided in American society, as a negative influence on creative development. On this basis Hypothesis No. 1 predicted that Creativity test performance scores of the Prince Edward County, Virginia, children would be significantly higher than the comparable scores of the Jackson, Michigan, children. The analysis shows there was no statistically significant difference, and thus the hypothesis is rejected. The differences that were found, although not significant, were in the opposite direction from that predicted. The mean total score for the Virginia children was 92.13 while for the Michigan students it was 97.9.

Table 5, page 64, depicts the significance of the differences in the comparative performances of Michigan and Virginia students at each of the four age levels and provides the test of the four sub-hypotheses, 1a through 1d. Hypothesis 1a had predicted no differences in performance between the six year old, or first grade children, in the two population samples. The rejection of Hypothesis 1a negates the important assumption that differences at other grade levels are primarily the consequence of differences in school experiences rather than differences in such factors as cultural backgrounds. The Michigan students were superior in their performance on Creativity tests at two of the other three grade levels, 6 and 8, but only for the former was the difference statistically significant.

When one examines the trend in performance scores as age increases, as reflected in Table 5, page 64, the Virginia sample shows progressively poorer comparative performance, with the exception of the oldest age level. This trend presents a temptation to conclude that school experience is a positive factor in creative performance.

The problem is that of determining how school experience produced the effect shown. Is it a direct result in the sense that the collective experiences of formal schooling enhance the mental capacities or processes that are primarily involved in creative expression, or at least creative behavior as it is elicited on the tests used in this study? Or, on the other hand, is this difference the result of secondary or indirect effects of formal schooling?

Basically there are two such secondary or indirect influences that merit consideration. The first of these might be called "facilitating factors." This set of factors can best be illustrated by pointing out that the Virginia children had several distinct handicaps which could not be controlled for, but which did affect their performance. In the first place, many of the Virginia children at grade levels 1, 4 and 6 displayed considerable difficulty in simply holding and manipulating a pencil. Indeed this seemed to increase with age across the three grade levels. Furthermore the task of labeling completed nonverbal stimuli impeded the Virginia children. In part this could be attributed to differences in vocabulary level and also to differences in the ability to spell, even approximately, those words used as labels. Both differences were a disadvantage to the Virginia children at all four grade levels. Third, if the student knew how to spell the appropriate label and proceeded on his own, the printing task was slow and laborious. It was observed that a large majority of the labels were printed at all levels in Virginia, whereas cursive writing was almost universal in the upper grade levels in Michigan.

The results from testing the four sub-hypotheses of the first hypothesis would tend to support this as a contributor to the between county differences. When these facilitating factors were equal at age 6, the Virginia students showed higher performance. This superiority is maintained at age level 9, but reduced considerably in magnitude. For the 11 year old children, where differences in the facilitating skills were the greatest, the superiority is reversed

and the degree of difference significant. If this trend were strictly a function of school experience on creative development, one would anticipate this difference to be maintained or increased at the 13 year old level. The performance differences are in the same direction, $\bar{x}_{m_{13}} > \bar{x}_{v_{13}}$, but the magnitude of difference is considerably reduced. The thirteen year old Virginia students were less handicapped in the areas of the facilitating skills than were their nine and eleven year old schoolmates.

The second indirect influence of school experience upon performance on tests of creative thinking is that of intelligence, as it is measured by standardized tests of intelligence. The data presented in the fifth chapter dealing with the relationship between I.Q. and Creativity scores can possibly add meaningful insight, as well as that data which analyzes the differential relationship between I.Q. and various Creativity Scale scores.

The overall correlation between total performance scores on the Creativity battery and the available I.Q. data was .21 which although low, is significant for the size of sample involved. Given this fact, and further, that this relationship is even higher for the so called "non-school" group, there would seem to be some justification for assuming that if the I.Q. data had been such that it could have been included in the analysis of variance model, and as such, controlled for, the resulting comparisons of "creative" thinking would have eliminated the difference between the two groups or possibly reversed that direction. It is highly doubtful, however, that such a reversal would have reached statistically significant proportions.

Figure 4, page 77, shows clearly that the overall superior performance scores of the Michigan children is primarily the consequence of superior performance on the Elaboration Scale. Figures 8 and 9, page 79, also show that this superiority was restricted to the 9 and

11 year old age levels. As was shown in Chapter 5, not only does the elaboration scale show by far the highest correlation with I.Q., it is at the age levels of 9 and 11 that the superiority of the Michigan children in I.Q. is most pronounced.

With these considerations tending to confound the meaning of the Michigan children's slightly superior performance on the creativity tests, the major conclusion to be drawn is that although formal schooling is not necessarily detrimental to the development of creative thinking abilities, in the situation covered by this study the evidence would indicate that formal schooling does little to directly and positively enhance creative abilities.

Although variations in amount of formal school experience did not produce statistically significant differences in creative performance scores, the factors of sex and age did. The ANOVA table, pages 61-62, shows that the main effect of age across both counties, sex and forms, and all scales was significant at the .001 level. This result is as predicted in Hypothesis #2: "The factor of age will make significant differences in the performance scores." There are several factors which must be considered in interpreting this result. In the first place, although there was a steady increase in total score across age groups, the differences between contiguous sample age groups was not significant between the age levels of 6 and 9, was significant but moderate between the ages of 9 and 11, and only between the 11 and 13 age levels is the difference decisive. Note that the first interval is three years and the other intervals two years. Thus in comparison to the types of gains with age typically found in terms of raw scores on intelligence or achievement tests, the reported creativity gains are moderate, if not minimal.

Figure 5, page 79, showing scale performance by age, clearly shows that the increase of performance scores with age are moderate and somewhat inconsistent for all the Scales except Elaboration which demonstrates a consistent and decisive increase with age. Again

the reader is reminded of the low correlation of elaboration with the other scales but the comparatively high correlation between I.Q. and Elaboration scores.

Finally, Figure 1, page 68, depicting verbal and nonverbal total scores across age, shows that the increase with age is true for nonverbal creativity, but by no means the case for verbal creativity. The latter shows a drop from age levels 6 to 9, a vertical plateau from ages 9 to 11, and a sharp increase from 11 to 13. One cannot help but conjecture as to the relationship of this observed phenomenon to the verbal emphasis of formal school experience. If one could substantiate the implications of this, Figure 1 might serve to prove the validity of the rationale that formal schooling is a suppressant of certain kinds of creative expression.

As has been indicated earlier, the limited study of the differential effect of sex on creativity has leaned towards boys as being somewhat superior. However, the research done in this area has been related to the effect of a given test stimulus. On the basis of the lack of any sound evidence to the contrary, Hypothesis #3 predicted: "There will be no significant differences in performance when compared by sex." However this hypothesis was rejected. The main effect of Sex was significant at the .05 level, with girls demonstrating superiority.

Figure 23, page 1 , Appendix C, shows that this difference cannot be explained by scores on the different tasks. The interaction of "Scale x Sex" was not significant indicating that this too would not be an adequate explanation of the superior performance of the girls. However this interaction effect approached significance and so bears further examination. The total Scale scores for boys and girls at each age level for both counties were compared and the resulting differences (plus or minus) were summed. The following differences were obtained. (A (+) in front of the figure indicates the

the sum showed a superiority for girls, a (-) a superiority for boys.)

| | |
|---|---|
| Fluency | + 2.3 |
| Flexibility | - 1.5 |
| Originality | - 2.2 |
| Elaboration | +27.0 |

Thus the girls superiority on total score was primarily a function of their superior performance on the Elaboration Scale.

Hypotheses 4, 5, 6a, 6b, and 7 are all concerned with differences in performance between the verbal and nonverbal forms of the creativity battery.

Hypotheses #4 had predicted: "The mean scores of the total study population will be significantly higher on the nonverbal forms in comparison to the verbal forms." The hypothesis was accepted on the basis of the main effect of Form being significant at the .005 level in the predicted direction. This result had been predicted because of the nonverbal characteristic of the populations sampled. This nonverbal characteristic was a function of both groups being basically comprised of educationally or socio-economically disadvantaged children and the fact that the age levels tested included children of ages, disadvantaged or not, when verbal skills are in relatively early stages of development or sophistication. Based on this latter factor, it was further predicted: (Hypothesis #5) "The differences between verbal and nonverbal scores will minimize with age."

This hypothesis was rejected. As Figure 1, page 68, shows, the differences increased with age. If this is further broken down by County it can be seen that the two samples show approximately parallel patterns, as evidenced by the lack of significance for the Form x County x Age interaction.

Thus it would seem that the rationale based on the nonverbal attributes of the sample population is not adequate. The addition of

another matching sample of "non-disadvantaged" children would have helped to clarify this issue. In hindsight a more appropriate derivation of hypotheses would have reflected the differential effect of formal school experience on verbal and nonverbal creative performance, with the rationale focusing upon the verbal emphasis of formal school experience.

However, it might be better to attribute this factor, or matrix of factors, that serve to suppress verbal creativity to a greater degree than nonverbal, to the general societal forces discussed in Chapter 2. This is supported by the rejection of sub-hypothesis 6a, "Virginia students will have higher scores on Nonverbal Creativity than the Michigan students," and the acceptance of sub-hypothesis 6b, "The Michigan students will have significantly higher scores on Verbal Creativity than the Virginia students." These results are summarized in Table 7, page 70.

Other results from the ANOVA table have either been referred to previously in interpreting the results of the testing of the hypothesis, or have little or no bearing on the major concern of this study; the effects of lack of formal school experience on creative thinking.

To summarize this section of the analysis, it can be stated that those children who have had normal school experience, in terms of number of years of continuous schooling, did better on tests of creativity than did those children whose schooling had been delayed or interrupted for varying amounts of time. Interpreting this result as implying that formal school experience encourages or supports the growth or development of those processes involved in creative thinking is of doubtful validity. First, the differences were not large, and not even significant for the overall population. Secondly, it has been shown that there is some degree of positive relationship between I.Q. and Creativity, and had this factor been statistically controlled, the results would have been more nearly equal.

Thirdly, it was indicated that one of the four scales, Elaboration, not only showed signs of being positively related to I.Q., and positively affected by school experience, but also was of questionable validity in the Creativity operational construct. This last claim is based on the fact that Elaboration generally showed equal or higher correlation with I.Q. than it did with the other Creativity scale scores. The bulk of the superiority of the Michigan students accrued from their performance on the Elaboration scale. Finally, certain skills highly related to school experience such as pencil mnnipulation, vocabulary, writing skills and spelling skills are utilized to varying degrees in the different assessment techniques. With respect to these skills, Virginia students were at an observable disadvantage. Again, it seems logical to assume that there would have been more equality in performance had the students been matched on these skills.

One general conclusions that would seem most warranted from the results of this study would be that the evidence indicates formal schooling has little, if any, direct positive effect on creative thinking, as measured by the tasks utilized.

One frustrating psychometric problem was reinforced by the results of the analysis of main effect of Scales. The main conclusion to be drawn is that of the interdependency of the scales and specifically of the dependency of Flexibility, Originality and Elaboration on Fluency. This interdependence has already been postulated as the reason for the effect of Scales being significant. As has been indicated in the previous chapter, this is viewed as a weakness in the construction of the creativity tasks and/or the mechanism of scoring these tests. It would seem that there are two possible ways of alleviating or eliminating this problem. One way would be to devise techniques which assess any one of the scales independently of the others. The second alternative would be the development of a formula

for deriving a Flexibility or Originality score that is in ratio to the obtained Fluency score. For example, given two individuals, one with a high Fluency score, the other with a low Fluency score, and both having an equal Originality score as derived by the scoring procedures utilized in this study, the individual with the lower Fluency score would end up with a higher Originality ratio or derived score. Further consideration of this problem is beyond the scope or concern of this paper.

The sections of data that dealt with problems involved in the measurement of the construct of Creativity supported some questions raised by other authors but also challenged some of these questions.

The analysis supported the assertions of Wallach and Kogan (1965) that using total task scores additively as indices of creativity as a construct discrete from intelligence is of questionable validity.

Evidence was found, however, that using scale scores summed across tasks does have construct validity. This assertion is strengthened if two additional factors are considered. The first of these is that this discreteness is more decisively illustrated in the sample population that is the most typical, the Michigan sample. The second factor is the questionable validity of the Elaboration scale, which alone showed a higher correlation with I.Q. than with the other scales.

Showing that the operational construct problem posited by Wallach and Kogan with respect to the intertask correlations does not hold true for the interscale correlations does not necessarily validate the use of a general creativity score. As had been discussed earlier, on pages 74 and 76, in a different context, the high interscore or interscale correlations, are in part a function of the dependence of Flexibility, Originality and Elaboration scores on Fluency scores and "creative" work is needed to produce assessment techniques or scoring procedures which minimize or eliminate this dependency. Given this consideration, a justification of a total score or general score by

combining the scale scores is premature, although in some ways certainly as defensible or justifiable as utilizing a "general intelligence" score.

Further argument against such general scores is provided by the correlations between the two forms of each scale score, verbal and nonverbal (page 104). This would suggest, at most, a total verbal and a total nonverbal creativity score. Undoubtedly the safest position to take to these psychometric considerations is the derivation and use of eight scores, both a verbal and nonverbal score for each scale.

However, the restricted nature of the sample, the testing conditions, and other factors so limit the degree to which inferences can be drawn, that little can be said in the way of conclusions other than that significant and weighty problems remain to be solved if much application is to be made of operational constructs of Creativity.

The final matter to be considered revolves around the differential effect by County on the relationship between I.Q. and Creativity. It should be recalled that I.Q. and Total Creativity Score showed a .36 correlation for the Virginia students and a .09 correlation for the Michigan students. Various explanations derived from the literature on the relationship between I.Q. and creativity measures prove inadequate for explaining these results. The following attempt to supply an interpretation or explanation is highly speculative. It is so both because of the limitations of the data itself and because it leaps beyond this data. However, the considerations involved seem of central importance to the issues of Creativity and those related issues dealing with the effect of education on creative development. With careful refinement any conclusions reached can be formulated into hypotheses and tested. This is as it should be, for the questions posed are, after all, empirical questions. Unfortunately, as will be pointed out, the chances for such empirical testing are, in fact, quite

limited. The value, if any, of the following theoretical discussion must lie not in the testable hypothesis it generates, but rather in the theoretical construct or framework for investigating the phenomena of creative behavior that are suggested. It seems highly possible to generate hypotheses that do test the efficacy of this framework.

Allegorically speaking, the theoretical rationale set the stage for a melodrama which depicted creativity as the charming and pure heroine and formal educational experiences as the dastardly villian. The insinuation made by the plot was that if one were to free the heroine from the clutches of the villian, even greater charm and purity would result. The role of the hero, the crucial helping hand, was daringly eliminated. It was also implied that our heroine was, in fact, confronted with two villians. The second being a black-hearted heavy, socialization, who lurked menacingly, although nebulously, in the shadows of the set.

Actually, two melodramas were involved, in the best psychological drama terms, one control and the other experimental. But alas and alack, the endings were the same.

The mean total creativity score, for all ages and both sexes, failed to show a significant difference, and in fact what difference there was ran contrary to what was hypothesized. How should this be interpreted? In what is somewhat of a face-saving device it has been suggested that although the study has failed to show up formal educational experiences as detrimental to creative development, it can be stated that such experiences have little positive effect, and with this revelation the villain was properly foiled again.

Attempts to pinpoint any differential effect in performance on these tests met with little success. When the data was broken down by Scale score or by Form, or both, and when the subjects were grouped by age or by sex, or both, as well as by "county", differences

were found here and there, but they failed to form anything approaching a pattern or trend which would support any interpretations. How much this consistency reflects an accurate picture of the behavioral phenomenon being investigated and how much it reflects error factors introduced by assessment techniques with at best marginal reliability, and testing conditions that were not ideal or standardized, is a question that cannot be answered. Moving from the creativity scores to I.Q. scores, the results of lack of formal schooling are, in contract, clearly evident. The Michigan students with schooling show a sharp increase over the first three to four grades of school and then a more gradual increase over the next four to five years. The Virginia students, without schooling, are equal to their Northern peers at age six but their scores dip sharply over the next five to six years. The scores of eighth grade students in Virginia were much higher than those of their fellow sixth graders; and, in fact, the former were nearly equal to their Michigan age mates. This last variation in the I.Q. pattern seems to be a result of these children having had three years of school prior to the closing of the Virginia schools. However, there may well be another explanation for the higher I.Q. level of the Virginia eighth graders. Attendance at the Freedom Schools from whose classes the Prince Edward County Sample was taken, was voluntary. It seems logical to assume that the proportion of all possible students that did elect to attend school decreases with age; and thus, at the older age level those who did attend were a more selective group on criteria such as school orientation or I.Q. than those in the lower age groups.

Given that the main effect of different school experience has little or no effect on one variable, creativity, but does have an effect on the other variable, I.Q., it does not necessarily follow that these differences in main effect will have a differential effect on the relationship between the two variables. The respective correlations for the two population groups could just as well be

equally high or equally low. Something more is needed to explain why one is moderately high and the other quite low. Nor can this be adequately explained on the basis of other factors frequently mentioned in supporting or criticizing the reported degree of relationship reported in various relevant studies. The three most prevalent ones can be considered here.

The first of these has to do with the variance, or lack of it, in one of the variables. Usually this has been focused on the data being derived from students in the higher ranges of I.Q., and thus the sample has an attenuated variance in I.Q. and generally a more normal variance for creativity. As sound as this argument is statistically, the writer has felt that it is still lacking construct-wise, for it does not explain why one variable would have such variance given the lack of variance on the other factor. However, above and beyond this concern, the argument of attenuated variation is not applicable to the data of this study. Although the variance in I.Q. distribution is substantially smaller than is the Creativity variance (range of I.Q. variances by grade level is 7.5-15.1, for Creativity 19.0-36.7), there is no substantial difference between the variance on either variable between the Virginia and Michigan sample. It would seem that in order to dismiss the results under discussion, the following relationship of variances would have to exist; higher I.Q. variance and lower Creativity variance for Virginia, and the reverse for the Michigan children. Although this is the direction the data shows, the between county variances on either I.Q. or Creativity were not substantial enough to explain the observed differences in the two "counties" on the correlation of I.Q. and Creativity scores.

As has been discussed previously, some writers have explained away the reported low relationships between I.Q. and Creativity on yet another statistical basis, the fact that when the sample group was of varied age levels, I.Q. scores used had a "built in" age control which the Creativity scores did not. One way of avoiding such

error would be to utilize raw scores, assuming that the scores would be derived from the same test. Another alternative is to group the subjects by age, compute the correlations with Creativity in these sub-groups, and compute the mean correlation. This is the procedure followed in this study since raw scores on I.Q. measures were not always available. The four age or grade levels utilized in the study's design represented a six month spread in ages at each level. Beyond this, the argument under consideration does not really apply, since it might have explained the low correlation for the Jackson children, had not the age factor been handled as it was, but it would in no way explain the substantial difference between these children and the Virginia children.

Another explanation of the nature of the relationship between I.Q. and Creativity is that of the threshold effect. In essence, this states that below a certain threshold point I.Q. and Creativity are significantly, if not even substantially, related; but beyond this threshold, frequently stated as 125-130 I.Q., the two variables have little or no relationship to one another. Beyond this point, higher "intelligence" is not necessarily, or usually, accompanied by higher creative ability, or higher creativity by higher "intelligence." The quotation marks used with the word intelligence are to indicate intelligence as measured by a standardized intelligence test. This explanation has merit, both in terms of its construct validity and in terms of the evidence that exists. However, notice that these effects have been limited to those efforts which locate the threshold within the I.Q. variable. One cannot help but wonder as to what type of results would be achieved by attempting to identify a threshold within the Creativity variable. The data of this study, with the relative low I.Q. for its subjects, does not lend itself to such an effort.

However, it appears to lend itself to a study of the threshold effect at the other end of the continuum, low I.Q. If the threshold

effect were to be symmetrical, then it should follow that the scores for the Michigan children more centrally located on the I.Q. distribution would show a higher correlation between I.Q. and Creativity than would the Virginia children whose I.Q. range is located further toward the extreme of the distribution. The results, of course, are just the opposite of what would be predicted by the symmetrical threshold effect. This should not be interpreted as a refutation of the threshold effect as postulated from the evidence of the other studies, but rather a statement showing that the threshold concept does not explain the phenomenon at hand, the substantially higher correlation between I.Q. and Creativity for the Virginia children.

Having found these explanations inadequate to explain the observed results, it is necessary, given also the limitations of this study, to step out on the proverbial limb to speculate or conjecture the causes for the differences between counties in the correlation between I.Q. and Creativity. It would be possible and perhaps even wise to dismiss this particular result by saying that the limitations of the data make inferences impossible. However, it seems necessary to make a brief attempt to explain the observed phenomena, not to produce conclusive statements, but in the twofold belief that: (a) the explanation that emerges has significance for the concerns of psychologyand education with the phenomena of creative behavior, and (b) the conclusions drawn represent hypotheses that are testable.

The creativity scores have been divided into various component parts, by scale, by verbal and nonverbal form, and even into combinations of these sub-units. When these various sub-total scores are correlated with I.Q., and these correlations are reported by "county," the overall impression is one of relatively parallel but not equivalent, sets of correlations. There are some exceptions, but speaking generally, the differences in correlations within a

sample group, e.g. Michigan, are less than the differences between the two major groups. Thus, this differential relationship between I.Q. and Creativity is relatively consistent across the various means of reporting creative performance and across the various age and sex sub-units of the population. What has not been done, and indeed cannot be done, at least in this study, and furthermore is seldom if ever done in similar studies, is a systematic breakdown of the I.Q. variables into component parts. The argument of the so-called "g" factor versus various discrete but related intelligence factors is an old dispute and although the weight of psychological concensus seemingly favors the discrete factors approach, we have not seriously deviated in our research or our pedagogy from the utilization of the global concept of intelligence.

"Intelligence" literature has ample discussion of the various factors influencing tests, e.g. numerical ability, vocabulary, reasoning, etc. In addition to these factors, generally determined a priori, an increasing number of studies have identified, but not always labeled, discrete factors through factor analysis or related techniques. A study of the relationship between these sub-factors and creativity and its sub-factors would undoubtedly disclose the existence of differential relationship. The author is not aware of any studies which have done this.

Conceivably the explanation for the results being considered could end here. The reasoning would go something like this. Assume the existence and definition of discrete intelligence factors which have differential degrees of relationship with creativity or the factors of creativity as suggested by the Scale constructs. In turn these factors are also influenced to different degrees by formal school experience. Furthermore, it must be that some of those factors, least dependent upon schooling for their acquisition, are more positively related to creative performance than other factors more dependent upon formal schooling. What is being suggested is, in a

way, a different threshold concept, or perhaps the same threshold effect described earlier. Given this dichotomy of intelligence factors, "school dependent" or "non-school dependent" one would assume that both groups of children tested for this study would show relative equivalence on the latter. Furthermore, if one were to compute the correlation of the sub-total of such factors with creativity indices, there would be equivalence between the groups as to the degree of this relationship. The observed differential on I.Q. for the two groups is the result of the additional I.Q. scores or points for the Michigan group derived from their successful performance on school dependent factors, points not achieved by the Virginia children. In addition, these dependent factors have a comparatively low degree of relationship with creativity. The resulting differences in I.Q. also serve to minimize the degree or correlation between I.Q. and creativity or, in another sense, maximize the differences in the degree of the relationship between the two groups. The threshold exists at that theoretical, or mystical, point which represents the end of the accumulation of scores from the "school dependent" items. All of this is based on tenuous, but testable, assumptions concerning the existence of I.Q. factors, which can be dichotomized as "school dependent" or "non-school dependent" and the further assumption that generally "school dependent" factors would have a lower correlation with creativity factors than would "non-school dependent" factors.

In essence the explanation is saying that the differential relationship between I.Q. and Creativity under consideration cannot be understood solely by quantitative differences in measured intelligence, qualitative differences must also be considered.

It would seem to the writer that qualitative differentiation of the general factor, creativity, has been attempted, although not completely or accurately accomplished, by the use of scale scores, and

also by the use of the verbal-nonverbal dichotomy. It is interesting to note that the differential relationship between I.Q. and creativity does not apply to all four scales, although it appears equal between verbal and nonverbal forms. Two of the scales, Flexibility and Originality, showed little difference between the two counties in the correlation between I.Q. and the Scale score. Flexibility was moderately correlated with I.Q. and Originality showed a low correlation. The difference between the two counties was confined to the correlation of I.Q. with Fluency and Elaboration. Fluency had moderate and significant correlation with I.Q., for the Virginia subjects, but a .00 correlation for the Michigan students. In Michigan, Elaboration showed a moderate correlation with I.Q., while in Virginia the relationship was substantial. As was also previously disclosed, the Virginia and Michigan children showed little difference in their performance on the various scales with the exception of Elaboration, where the Michigan children showed clear superiority. Being unable to be specific concerning the interrelationships between the speculative qualitative dimensions of intelligence and of creativity, these differences on performance could be explained, within the proposed context, in the following manner. Fluency would seem moderately correlated with the "non-school dependent" dimensions which both groups share, but seems to show little relationship with the "school dependent dimensions." Thus, both perform equally and this is reflected in the moderate correlation with I.Q. for the Virginia students when their scores are primarily derived from the "non-school dependent" dimension. For the Michigan children, the I.Q. score reflects the additional "school dependent" factors which do not serve to increase Fluency performance. Therefore, the resulting correlation is diminished. Elaboration, on the other hand, while substantially correlated to the "non-school dependent" factors is also highly related to certain "school dependent" factors and this is reflected in the higher performance by Michigan children. At the same time the correlation of I.Q. with Elaboration score is high compared to other scales, it is still higher for Virginia students than for Michigan students,

reflecting a low correlation with other "school dependent" factors. The relative performance on the other two scales could also be explained within this context, but the above can serve as an adequate example.

The above explanation of the differential relationship between I.Q. and Creativity is most tenuous, if not feeble. Its merit is solely that it is testable. It also serves to illustrate the need for qualitative, as well as quantitative analysis of such psychological constructs as creativity and intelligence.

As important as present and future efforts in this direction are, it would seem that their ultimate productivity is dependent upon the development of a more general construct of intelligence.

It is something of a paradox that what ultimately evolves from a projection of this rationale is both an attack on the use of a general intelligence score and a plea for the emergence of a more general concept of intelligence. Of course, the term intelligence is being used in two different ways. A general intelligence score refers to intelligence as it is measured by some standardized assessment technique, while the latter usage refers to intelligence in the sense of "intelligent" behavior, for which the primary criterion is whether or not the behavior is adaptive to one's environment.

Again, for the sake of brevity, over-simplification characterizes this discussion. If intelligence were differentiated according to qualitative factors, and the same were true of creativity, it would seem more meaningful to group them together, undoubtedly along with other factors not implied by the popular notion of either I.Q. or creativity. They are all dimensions of intelligent (adaptive) behavior. The present dilemma, where by logical deduction can lead to defining creativity as non-intelligent behavior, is hardly healthy. Nor does it seem to the author that continued effort to solidify the

construct of creativity, with the problems engendered by both the diversity of existing definitions and their nebulous quality, seem promising. A more productive path would seem to be the identifying and defining along qualitative lines, those mental processes involved in intelligent behavior. Given the concomitant development of appropriate assessment techniques, the various interrelationships between the factors individually or in combination, the effect of various experiences, school oriented or otherwise, on the development and use of these capacities, and other empirical considerations can be systematically investigated. Given a construct of this nature, it would seem that the measure of intelligence could be more meaningfully viewed as a means of assessing the appropriateness or efficiency with which the individual factors or combinations of factors are utilized in given situations. Creativity then might more appropriately be perceived as a style or predisposition towards certain identified patterns of applying or utilizing these processes, rather than as a separate process or set of processes.

## Bibliography


Anderson, H. H., "Creativity and Education", College and University Bulletin, Vol. 13, No. 14, May 1, 1961.

Barron, F., "The Disposition Toward Originality", The Journal of Abnormal and Social Psychology, Vol. 51, November 1955.

Barron, F., "The Psychology of Imagination", Scientific American, 1958, p. 199.

Barzun, J., The House of Intellect, New York: Harper and Brother, 1959.

Biber, Barbara, "Premature Structuring as a Deterrent to Creativity, The Teacher's Role in Creativity, A Symposium, 1958", American Journal of Orthopsychiatry, Vol. XXIX, No. 2, April 1959.

Bronfenbrenner, U., "The Changing American Child -- A Speculative Analysis", The Journal of Social Issues, Vol. 17, No. 1, 1961.

Drews, Elizabeth M., "A Critical Evaluation of Approaches to the Identification of Gifted Students", American Council on Education, Washington, D. C., 1961.

Getzels, J. W., "Non-I.Q. Intellectual and Other Factors in College Admission", The Coming Crisis in the Selection of Students for College Entrance, American Educational Research Association, Washington, D. C., 1960.

Getzels, J. W., and Jackson, P. W., "The Social Context of Giftedness: A Multi-Dimensional Approach to Definition and Methods", E. P. Torrance (ed.), Creativity: Second Minnesota Conference on Gifted Children, Minneapolis, Minnesota Center for Continuation Study, University of Minnesota, 1960.

Getzels, J. W., and Jackson, P. W., Creativity and Intelligence, New York: John Wiley and Sons, Inc., 1962.

Getzels, J. W., "Creative Thinking, Problem-Solving, and Instruction", E. Hilgard (ed.), Theories of Learning and Instruction: The Sixty-Third Yearbook of the National Society for the Study of Education Part 1, Chicago, Illinois, University of Chicago Press, 1964.

Ghiselin, B. (ed.), The Creative Process, New York: Mentor Books, 1952.

Golann, E., "Psychological Study of Creativity", Psychological Bulletin, Vol. 60, No. 6, November 1963, pp. 548-565.

Greenhouse, S. W., and Geisser, S., "On Methods in the Analysis of Profile Data", Psychometrika, 1959, 24, pp. 95-112.

Grover, B. L., "Some Effects and Correlates of Different Types of Practices Used in Studying a Topic in Ninth Grade Classrooms", Unpublished Doctoral Dissertation, University of Minnesota, 1963.

Guilford, J. P., Kettner, N. W., and Christensen, P. R., "A Factor-Analytical Study Across the Domains of Reasoning, Creativity, and Evaluation: I. Hypotheses and Description of Tests", University of Southern California Psychological Laboratory Reports, No. 11, 1954.

Guilford, J. P., Kettner, N. W., and Christensen, P. R., "A Factor-Analytic Study Across the Domains of Reasoning, Creativity, and Evaluation: II. Administration of Tests and Analysis of Results", University of Southern California Psychological Laboratory Reports, No. 16, 1956.

Guilford, J. P., "The Relation of Intellectual Factors to Creative Thinking in Science", C. Taylor (ed.), The 1955 University of Utah Research Conference on the Identification of Creative Scientific Talent, Salt Lake City, University of Utah Press, 1956, pp. 69-95.

Guilford, J. P., "Creative Abilities in the Arts", Psychological Review, Vol. 64, 1957, pp. 110-118.

Guilford, J. P., "Three Faces of Intellect", American Psychologist, Vol. 14, 1959a, pp. 469-479.

Guilford, J. P., "Intellectual Resources and Their Values as Seen by Scientists", C. Taylor (ed.), The 1959 University of Utah Research Conference on the Identification of Creative Scientific Talent, Salt Lake City, University of Utah Press, 1959b, pp. 128-149.

Guilford, J. P., "Traits of Creativity", H. Anderson (ed.), Creativity and its Cultivation, New York: Harper and Brother, 1959c, pp. 142-161.

Henry, J., "The Problem of Spontaneity, Initiative and Creativity in Suburban Classrooms, The Teacher's Role in Creativity -- Symposium, 1958", American Journal of Orthopsychiatry, Vol. XXIX, No. 2, April 1959, pp. 266-279.

Hoffman, B., The Tyranny of Multiple-Choice Tests, New York: Harper and Brother, March 1961, pp. 222, 37-44.

Jex, F., "Negative Validities for Two Different Ingenuity Tests", C. W. Taylor (ed.), The Third (1959) University of Utah Research Conference on the Identification of Creative Scientific Talent, Salt Lake City, University of Utah Press, 1959.

LaBrant, L., "The Dynamics of Education", Saturday Review, 1959, 42 (37).

Luchins, A., "On Some Aspects of the Creativity Problem in Thinking", Harmes (ed.), Fundamentals of Psychology: The Psychology of Thinking -- Annals of the New York Academy of Sciences, Vol. 19, Art. 1, December 23, 1960.

MacKinnon, D. W., "The Nature and Nurture of Creative Talent", American Psychologist, 1962, p. 17.

McNeil, E. B., "The Paradox of Education for the Gifted", Improving College and University Teaching, 8, (3), 1960, pp. 111-115.

Morgan, E. S., "What Every Yale Freshman Should Know", Saturday Review, 43 (4), 1960.

Ripple, R. C., and May, F. B., "Caution in Comparing Creativity and I. Q.", Psychological Reports, 1962, 10, pp. 229-230.

Rugg, H., Imagination, New York: Harper and Row, 1963.

Schactel, E. G., Metamorphosis, New York: Basic Books, 1959.

Schenitzki, D. P., "Adult Evaluation and Peer Evaluation as Factors in Creative Thinking", E. P. Torrance (ed.), New Educational Ideas: Proceeding of the Third Minnesota Conference on Gifted Children, Minneapolis Center for Continuation Study, University of Minnesota, 1961.

Smillie, D., "Tests and Definitions of Intelligence", (Mimeograph), Merrill-Palmer Schools, Detroit, 1959.

Spencer, L., "Implementation -- The Effective Use of Identification Techniques", J. B. Conant (ed.), The Identification and Education of the Academically Talented Student in the American Secondary School, Washington, D. C., National Education Association, 1958.

Strong, Ruth, "Developing Creative Powers of Gifted Children", Creativity of Gifted and Talented Children, New York: Bureau of Publications, Teacher College, Columbia University, 1959.

Taylor, C. W. (ed.), The Third (1959) University of Utah Research Conference on the Identification of Creative Scientific Talent, Salt Lake City, University of Utah Press, 1959.

Taylor, C., "Identifying the Creative Individual", E. P. Torrance (ed.), Creativity: Second Minnesota Conference on Gifted Children, Minneapolis Center for Continuation Study, University of Minnesota, 1960.

Taylor, I. A., "The Nature of the Creative Process", P. Smith (ed.), Creativity, New York: Hastings House, 1959, pp. 51-82.

Torrance, E. P., Guiding Creative Talent, Englewood Cliffs, New Jersey: Prentice Hall, Inc., 1962a.

Torrance, E. P., "Conditions for Creative Learning", Association For Childhood Education International, April 1963a.

Torrance, E. P., Education and the Creative Potential, Minneapolis, The University of Minnesota Press, 1963b.

Torrance, E. P., "Changing Reactions of Preadolescent Girls to Tasks Requiring Creative Thinking", The Journal of Genetic Psychology, 1963c.

Torrance, E. P., Administration and Scoring Manual for Abbreviated Form III, Minneapolis Tests of Creative Thinking, University of Minnesota, Minneapolis, Minnesota, Bureau of Educational Research, 1963d.

Torrance, E. P. and Gowan, J. C., "The Reliability of the Minnesota Tests of Creative Thinking", Bureau of Educational Research, College of Education, University of Minnesota, BER-63-4, August, 1963e.

Tumin, M., "Obstacles to Creativity", ETC: A Review of General Semantics, 1953, 11, pp. 261-271.

Vernon, P. E., Intelligence and Attainment Tests, London: University of London Press, 1960.

Wallach, M. A., and Kogan, N., Modes of Thinking in Young Children, Chicago: Holt, Rinehart and Winston, 1965.

Wolfle, O., "Diversity of Talent", American Psychologist, 1960, 15, pp. 535-545.

Whyte, W. H., The Organization Man, New York: Simon and Schuster, Inc., 1956.

Winer, B. J., Statistical Principles in Experimental Design, New York: McGraw-Hill, 1962.

Yamamoto, K., "Creativity and Sociometric Choice Among Adolescents", Unpublished Master's thesis, University of Minnesota, 1960.

## Appendix A

### Results of the reliability study

As discussed in Chapter 4, a test-retest reliability study, with an approximate twelve month interval was conducted on a major portion of the original Prince Edward County study on the two nonverbal tasks of Figure Completion and Circles.

The reliability data is reported in Table 22. This table presents the data for each of the four scales on each of the two tasks, plus the nonverbal sub-total scale score, total task score for the two tests, and the total nonverbal score. The data is also differentiated by sex and by age level, and then the sexes are combined to present total age level reliabilities and, the reliabilities for the total sample.

Figures 17 and 18 illustrate the differential effect of sex and age on the obtained coefficients of reliability. Both figures report the nonverbal scale reliabilities and nonverbal total score reliability.

Table 22

Test-retest reliabilities obtained on nonverbal tests of Minnesota Tests of Creative Thinking, Abbreviated Form VII

| | Fluency | | | Flexibility | | | Originality | | | Elaboration | | | Total Score | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | FC | C | NV ST | FC | C | NV ST | FC | C | NV ST | FC | C | NV ST | FC | C | NV ST |
| **FEMALE** | | | | | | | | | | | | | | | |
| A.L. 6(15) | .31 | -.14 | .03 | .16 | .23 | .14 | -.17 | .68 | .32 | .04 | .05 | -.16 | .14 | .16 | .12 |
| A.L. 9(10) | .61 | .76 | .86 | .28 | .42 | .61 | -.08 | .63 | .44 | .39 | .44 | .35 | -.40 | .48 | .31 |
| A.L. 11(10) | .45 | .82 | .77 | .54 | .52 | .49 | -.18 | .54 | .49 | .42 | .40 | .55 | .39 | .59 | .50 |
| A.L. 13( 7) | X | .73 | .74 | .76 | .78 | .84 | .25 | .53 | .63 | .79 | .90 | .84 | .72 | .68 | .72 |
| T.S. (42) | .44 | .54 | .56 | .42 | .59 | .59 | -.08 | .52 | .44 | .63 | .52 | .65 | .42 | .62 | .62 |
| **MALE** | | | | | | | | | | | | | | | |
| | Fluency | | | Flexibility | | | Originality | | | Elaboration | | | Total Score | | |
| | FC | C | NV ST | FC | C | NV ST | FC | C | NV ST | FC | C | NV ST | FC | C | NV ST |
| A.L. 6(10) | X | .06 | .03 | .67 | .57 | .64 | .25 | .40 | .27 | .61 | .58 | .86 | .52 | .48 | .59 |
| A.L. 9(10) | -.41 | .76 | .72 | .08 | .70 | .65 | -.14 | .52 | .39 | .30 | .73 | .80 | .10 | .73 | .55 |
| A.L. 11( 7) | .93 | .22 | .31 | -.26 | .05 | .02 | .68 | .58 | .85 | .04 | .33 | .28 | .74 | .80 | .92 |
| A.L. 13( 8) | .52 | .72 | .84 | .42 | .72 | .71 | .58 | .25 | .65 | .10 | .60 | .53 | .73 | .54 | .89 |
| T.S. (35) | .82 | .60 | .71 | .64 | .64 | .73 | .47 | .40 | .52 | .53 | .67 | .78 | .67 | .70 | .77 |
| **MALE-FEMALE COMBINED** | | | | | | | | | | | | | | | |
| | Fluency | | | Flexibility | | | Originality | | | Elaboration | | | Total Score | | |
| | FC | C | NV ST | FC | C | NV ST | FC | C | NV ST | FC | C | NV ST | FC | C | NV ST |
| A.L. 6(24) | .46 | .07 | .04 | .41 | .34 | .33 | -.17 | .42 | .27 | .78 | .19 | .25 | .17 | .31 | .29 |
| A.L. 9(20) | .26 | .77 | .73 | .30 | .59 | .60 | -.09 | .47 | .38 | .28 | .54 | .57 | -.07 | .64 | .45 |
| A.L. 11(17) | .55 | .29 | .50 | .30 | .26 | .32 | .15 | .34 | .45 | .39 | .26 | .40 | .43 | .36 | .64 |
| A.L. 13(15) | .71 | .74 | .74 | .60 | .85 | .84 | .48 | .51 | .63 | .70 | .78 | .85 | .72 | .71 | .85 |
| T.S. (77) | .65 | .54 | .60 | .48 | .62 | .67 | .17 | .45 | .48 | .59 | .59 | .68 | .53 | .66 | .69 |

Key to abbreviations:
- FC - Figure Completion Test
- C - Circles Test
- NVST - Nonverbal Sub-total Score obtained by adding scores for the two nonverbal tests
- A.L. - Age level
- ( ) - Number in ( ) is the N for each sample group
- T.S. - Total sample across age levels
- X - Indicates those sample groups where scores were identical for all subjects on both testing sessions. In this case all subjects had Fluency scores of 10 in both testing situations.

Figure 17

Sex differences in obtained coefficients of reliability for total scale scores & total score of the nonverbal tasks of the MINNESOTA TESTS OF CREATIVE THINKING, ABBREVIATED FORM VII

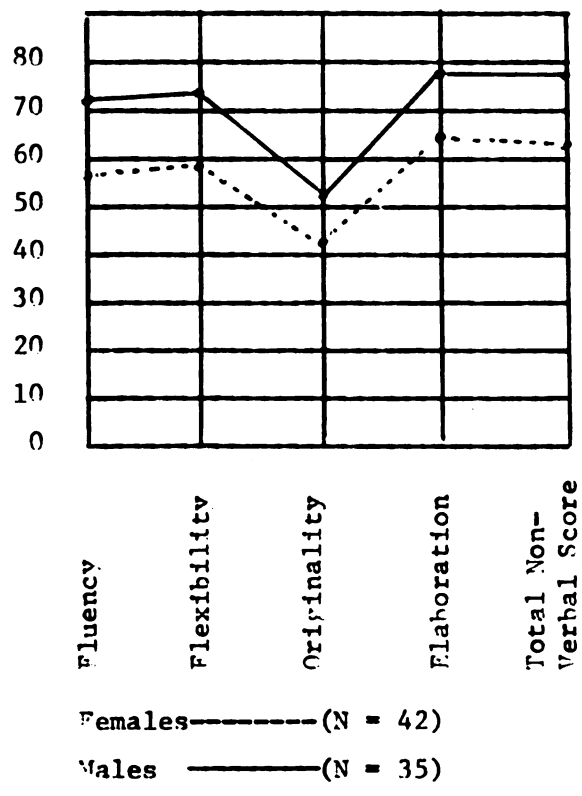
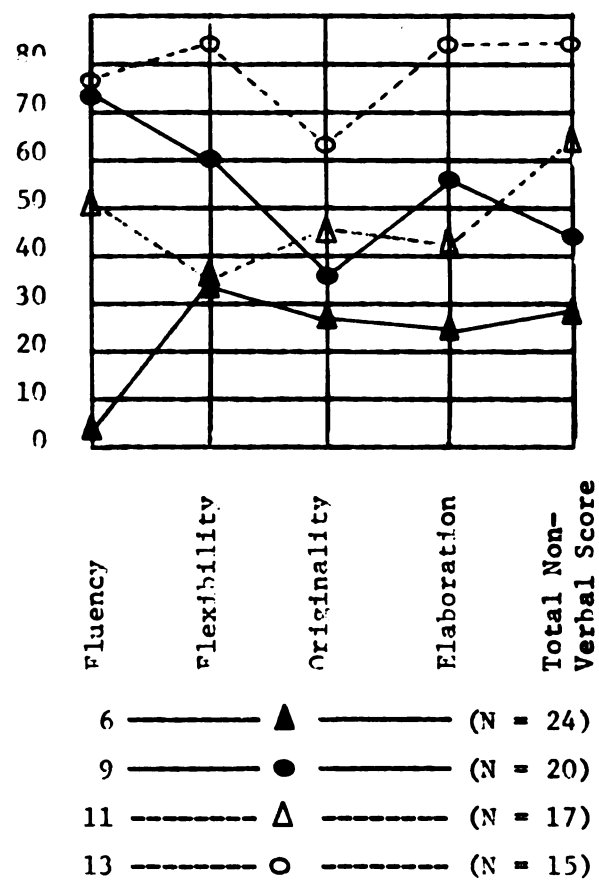



Figure 18

Age differences in obtained coefficients of reliability for total scale scores and total score of the nonverbal tasks of the MINNESOTA TESTS OF CREATIVE THINKING, ABBREVIATED FORM VII



## Appendix B

This section illustrates the procedure developed to explain the three-way interaction of "County x Sex x Age". This interaction effect was a significant source of variance, as had been the main effects of "age" and "sex". However the main effect of "County" as well as all the two way interactions "County x Sex", "County x Age", and "Sex x Age" were not identified as significant. Neither Figure 3, page 73, or an analysis of the data using the Neuman-Keuls procedure gave evidence of factor(s) causing the significant interaction effect. Thus the following procedure was employed. This procedure establishes a sequence of expected values derived from the results of the analysis of main effects and higher order interactions and establishes the extent of deviation of the observed mean scores from the derived or expected mean scores. The technique can best be understood by following through the procedure as described below.

The observed mean for individuals across all groups is 94.5. Given this and the condition that the main effect of county was not significant the expected mean for both Virginia and Michigan students would be 94.5. The main effect of Sex was significant and the mean for boys was approximately 90 and for girls, 99. Thus, the model to this point would show:

Expected mean 94.5

Boys $\underline{-4.5}$ = 90

Girls $\underline{+4.5}$ = 99.0

The main effect of age was also significant with the following observed means for the four age levels.

Age 6 = 83.0 (94.5) + ($\underline{-11.5}$) = 83.0

Age 9 = 87.0 (94.5) + (- 7.5) = 87.0

Age 11 = 94.5 (94.5) + ( 0 ) = 94.5

Age 13 = 113.5 (94.5) + (+19.0) = 113.5

Combining these effects we derive the following set of expected means by age and by sex. These are recorded in Table 23.

Table 23

Expected means by age by sex

| County Sex Age | |
|---|---|
| Girls ( 6) 94.5 + 4.5 - 11.5 = | 87.5 |
| Girls ( 9) 94.5 + 4.5 - 7.5 = | 91.5 |
| Girls (11) 94.5 + 4.5 + 0 = | 99.0 |
| Girls (13) 94.5 + 4.5 + 19 = | 118.0 |
| Boys ( 6) 94.5 - 4.5 - 11.5 = | 78.5 |
| Boys ( 9) 94.5 - 4.5 - 7.5 = | 82.5 |
| Boys (11) 94.5 - 4.5 + 0 = | 90.0 |
| Boys (13) 94.5 - 4.5 + 19 = | 109.0 |

Table 24 contains the deviation values and their derivation by subtracting the expected mean values from Table 23 from the observed mean values from Table 3 (pp.59-60). The error term derived from the analysis of variance is roughly 9 so the deviations in parentheses are those equal to or less than the magnitude of the error term, or those which do not exceed the probabilities of chance occurance. The remaining deviations would be the primary sources of the obtained

Table 24

Deviation of observed means from expected means with data grouped by age, sex and county

| | Age 6 | Age 9 | Age 11 | Age 13 |
|---|---|---|---|---|
| Virginia boys | 89.0<br>-78.5<br>10.5 | 85.0<br>-82.5<br>(2.5) | 78.0<br>-90.0<br>-12.0 | 86.0<br>-109.0<br>- 23.0 |
| Virginia girls | 86.0<br>-87.5<br>(-1.5) | 90.0<br>-91.5<br>(-1.5) | 90.0<br>-99.0<br>(-9.0) | 133.0<br>-118.0<br>15.0 |
| Michigan boys | 69.0<br>-78.5<br>(-9.5) | 76.0<br>-82.5<br>(-6.5) | 104.0<br>-90.0<br>14.0 | 133.0<br>-109.0<br>24.0 |
| Michigan girls | 88.5<br>-87.5<br>(1.0) | 97.0<br>-91.5<br>(5.5) | 106.0<br>-99.0<br>(7.0) | 134.0<br>-118.0<br>- 16.0 |

differences. The across age performance trend of boys and the across county and sex performance of the thirteen year olds would seem to be making contributions to the significance of the three way interaction. This table also shows that "Sex x Age" was not significant because the two groups in both counties compensated each other. Similarly, "County x Sex" failed to be significant because the younger children compensated for the older or vice versa in each of the four groups. Only "County x Age" does not show such a complete or consistent pattern of compensatory factors. Table 4 shows this to have been significant at the .10 level. Although this is not interpreted as significant, if this two way interaction is integrated into the model of expected values, the identification of the sources of significance for the three way interaction of "Sex x County x Age" becomes more specific. The means of the four age groups for each county are as follows:

Virginia - Age Six = 87.5 (94.5 - $\frac{7}{}$ = 87.5)

Michigan - Age Six = 79.0 (94.5 - $\frac{15.5}{}$ = 79.0)

Virginia - Age Nine = 87.5 (94.5 - 7 = 87.5)

Michigan - Age Nine = 86.5 (94.5 - 8 = 86.5)

Virginia - Age Eleven = 84.0 (94.5 - 10.5 = 84.0)

Michigan - Age Eleven = 105.0 (94.5 + 10.5 = 105.0)

Virginia - Age Thirteen= 109.0 (94.5 + 15.0 = 109.5)

Michigan - Age Thirteen= 117.5 (94.5 + 23.0 = 117.5)

The subsequent set of expected values and their derivations are given in Table 25.

Table 25

Table for derivation of expected mean values for data grouped by county by sex by age

| | Mean | Effect C x A | Effect Sex | Effect Age | Expected Mean |
|---|---|---|---|---|---|
| Virginia - 6-F | 94.5 | - 7 | +4.5 | -11.5 | 80.5 |
| Michigan - 6-F | 94.5 | -15.5 | +4.5 | -11.5 | 82 |
| Virginia - 6-M | 94.5 | - 7 | -4.5 | -11.5 | 71.5 |
| Michigan - 6-M | 94.5 | -15.5 | -4.5 | -11.5 | 63.0 |
| Virginia - 9-F | 94.5 | - 7 | +4.5 | - 7.5 | 84.5 |
| Michigan - 9-F | 94.5 | - 8 | +4.5 | - 7.5 | 83.5 |
| Virginia - 9-M | 94.5 | - 7 | -4.5 | - 7.5 | 75.5 |
| Michigan - 9-M | 94.5 | - 8 | -4.5 | - 7.5 | 74.5 |
| Virginia - 11-F | 94.5 | -10.5 | +4.5 | 0 | 88.5 |
| Michigan - 11-F | 94.5 | +10.5 | +4.5 | 0 | 109.5 |
| Virginia - 11-M | 94.5 | -10.5 | -4.5 | 0 | 79.5 |
| Michigan - 11-M | 94.5 | +10.5 | -4.5 | 0 | 100.5 |
| Virginia - 13-F | 94.5 | +15.0 | +4.5 | +19.0 | 133.0 |
| Michigan - 13-F | 94.5 | +23.0 | +4.5 | +19.0 | 141 |
| Virginia - 13-M | 94.5 | +15 | -4.5 | +19.0 | 124 |
| Michigan - 13-M | 94.5 | +23 | -4.5 | +19.0 | 132 |

Table 26 contains the deviations of the observed values from the expected values when the data is grouped according to Age and Sex and County. The derivation scores shown were obtained in the same manner as illustrated in Table 24, page 136, but only the derivation values are shown in Table 26.

Table 26

Deviation of observed means from expected means for data grouped by county by age by sex
( ) Deviations equal to or lower than error effect

| | Age 6 | Age 9 | Age 11 | Age 13 |
|---|---|---|---|---|
| Virginia Boys | 17.5 | (9.5) | (-1 ) | -38 |
| Virginia Girls | (5.5) | (6 ) | ( 1 ) | (-5 ) |
| Michigan Boys | (6 ) | (1.5) | ( 3.5) | ( 1.5) |
| Michigan Girls | (7 ) | 13.5 | (-3.5) | -39 |

When the error effect of approximately 9.0 is used to eliminate deviations that may be a function of error, the number of significant deviations of the observed means from the derived expected mean values is reduced to four specific subgroup means. There are those of the Virginia boys at ages six and thirteen and those of the Michigan girls at ages nine and thirteen which are clearly contributors to the interaction effect. Table 25 reflected several significant deviations that did not show up as significant on Table 26 when the "County x Age" interaction affect was included. These were the Virginia boys, age 11, and Virginia girls, age 13, and the Michigan boys at ages 11 and 13. The fact that these are minimized by adjusting the expected mean to allow for the "County x Age" interaction would identify these as the major contributors to the near significance of that interaction effect rather than the three way interaction presently being considered.

Thus the significance of the three way interaction "County x Sex x Age" seems to be a function of having two of the four major sample groupings with one age level's mean observed performance score considerably below the mean for the respective group that is obtained from a sequence of establishing expected means based on the results of the analysis of variance model. These two groups are both at the thirteen year old age level and are Virginia Boys (-38) and the Michigan Girls (-39). Although the age level is the same, the sex is different in the two counties. This same process of analysis identified an age level in each sample group whose performance scores were above the derived expected mean values, six year old Virginia Boys (+17.5) and nine year old Michigan Girls (+13.5). Thus in this case not only sex but age level is different between the two counties.

Thus the differential performance of the Virginia Boys, age six and thirteen and the Michigan Girls, ages nine and thirteen, contribute the sources of variance that result in the "County x Sex x Age" interaction being a source of significant difference.

## Appendix C

### Analysis of creativity task scores

Appendix C contains the graphs illustrating the comparative performance on the four tasks that collectively comprise the Abbreviated Form VII, Minnesota Tests of Creativity. The prior analysis of variance ignored this reference, focusing upon a total score, or this total score broken down into scales (Fluency, Flexibility, Originality and Elaboration) or form (Verbal and Non-verbal). Figures 19 through 22 show comparative performances on the four tasks in each of the four grade levels for the sample grouped by county (Virginia, Michigan) and by sex.

When the data are graphed to show scores on tasks, Figures 19 through 22 again demonstrate perhaps the most consistent features of the data being analyzed, and that is its lack of consistency. The performance levels for the four groups tend to disperse themselves in a manner that makes explanation difficult and tenuous. The question must be raised as to whether this condition is more a function of error term introduced as a consequence of task

Figure 19

Age group mean scores on Incomplete Figures Task for four subgroups (County x Sex)

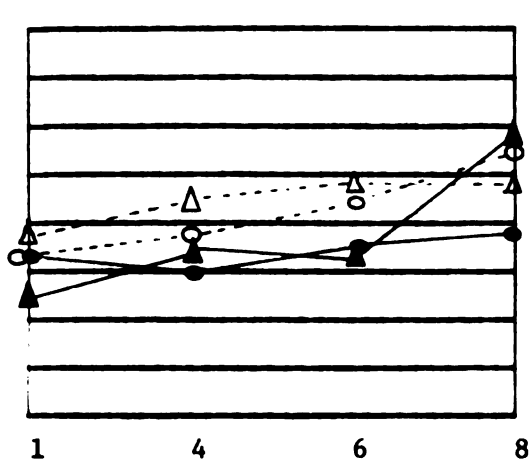



Figure 20

Age group mean scores on Circles Task for four subgroups (County x Sex)

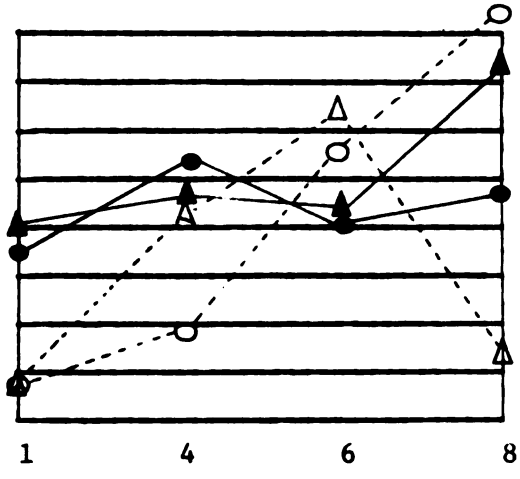



Figure 21

Age group mean scores on Product Improvement Task for four subgroups (County x Sex)

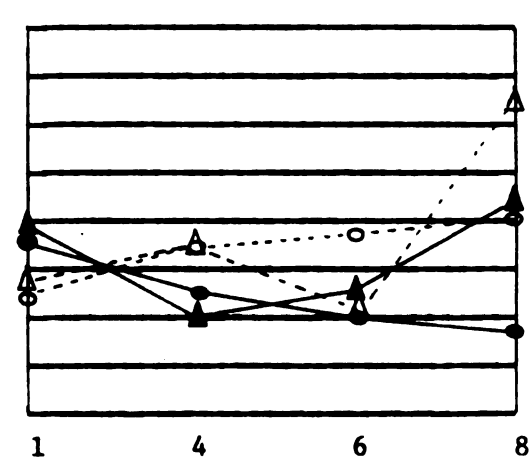
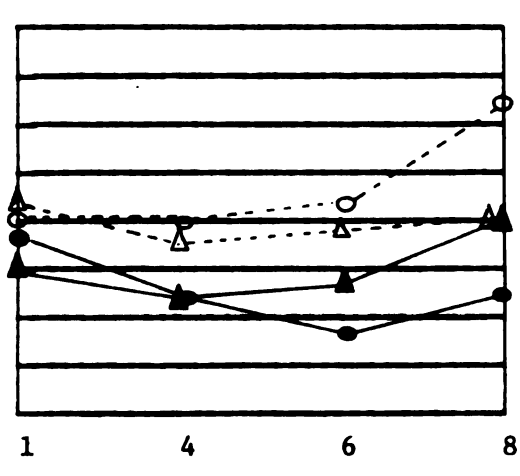



Figure 22

Age group mean scores on Uses Task for four subgroups (County x Sex)




Michigan Male
Michigan Female
Virginia Male
Virginia Female

construction or the testing situations, or a combination of both, rather than a function of the characteristic of cognitive styles or creative thinking abilities indigenous to the samples as grouped for the graphs.

The Michigan girls are a good example of this effect. On two of the tasks they show a relative plateau effect (Incomplete Figures and Uses), while on one (Product Improvement) they show a relative plateau effect through the first three grade levels and then an extremely sharp increase at age 13. The other task shows no plateau, only a sharp increase over the first three levels and then a sharp decrease, or perhaps plunge would be a better word, at the oldest age level. It is difficult not to assume that the latter phenomena, a sharp drop at age 13, represents some abnormality in the testing procedure. It cannot be explained on the basis of age since the other three subgroups showed increases at this age, in fact, two of them showed sharp increase. It cannot be explained either on the basis of county, since the Michigan males showed a sharp increase, and finally it cannot be explained on the basis of sex, since the Virginia girls had the sharpest increase of any of the groups at the final grade level.

As the purpose of this section of the data is to enhance the major analysis, no further time will be spent on this secondary analysis. The above example should suffice to illustrate the characteristic of inconsistency.

There are several other ways of presenting these data that demonstrate possible task effects that might influence the analysis of the data when the individual task scores are ignored.

The first of these descriptions shows the task performance differentiated by county.

Figure 23

Task performance differentiated by county

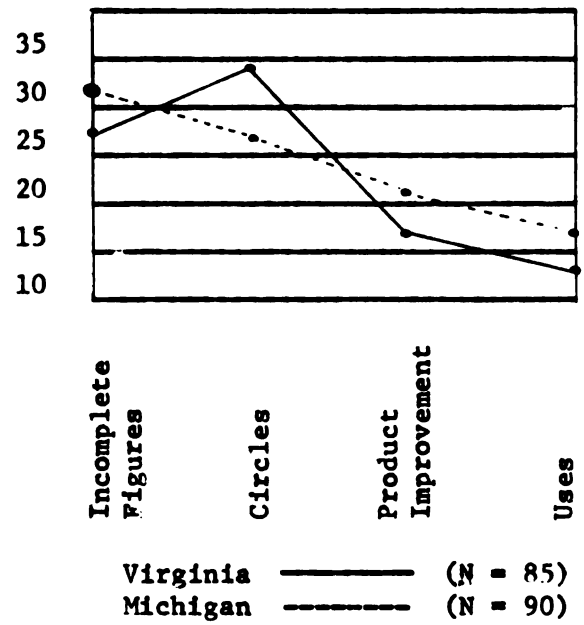



Figure 23, depicting the scores on the individual tasks for the sample divided by County shows somewhat more consistency. The Michigan children's performance was superior to the Virginia students on three of the four tasks with the difference amounting to four points on the two verbal tasks and three points on Figure Completion. The Virginia students scores were higher than their Michigan peers by seven points on the Circles task. Undoubtedly this latter result reflects the poor performance by the thirteen year old Michigan girls, which, as earlier indicated, is suspect of testing irregularities.

Figure 24 groups the scores by Counties together, but separates the data into the four age groups and compares their performance on the individual tasks.

Figure 24

Scores on creativity tasks for each of four age groups

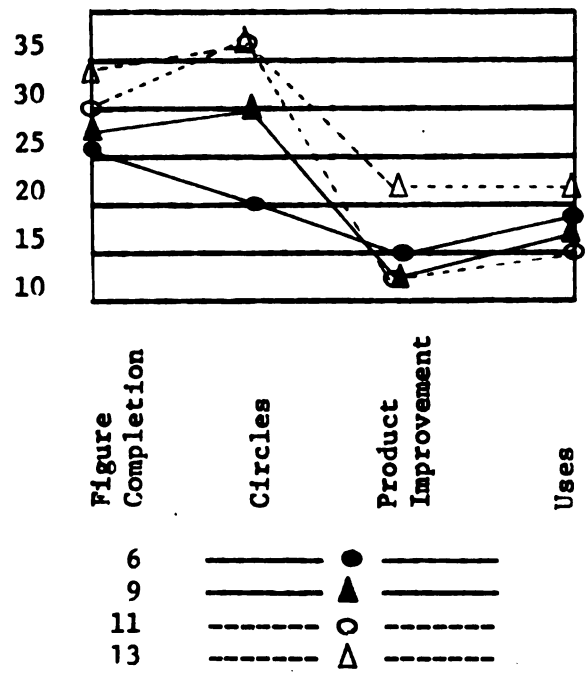
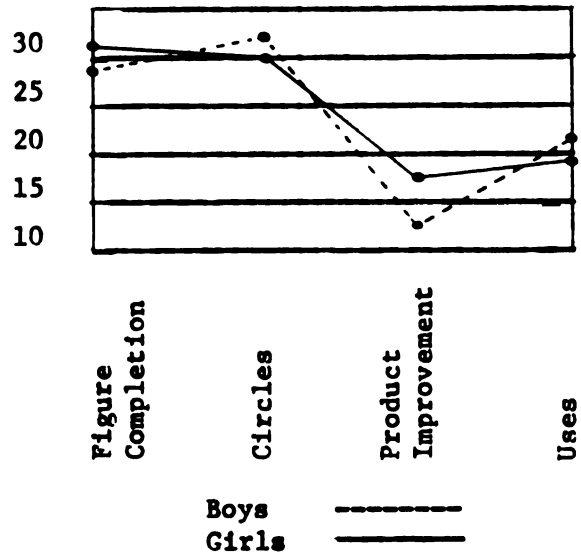



Figure 24 simply adds emphasis to the phenomenon described previously in Figure 1, page 68, depicting form performance by age. It adds the insight that the "valley" effect at grade levels four and six on verbal creativity was a function of both tasks; in fact the mean scores for the two grade levels were almost equal on the two tasks. The steady increase with age for nonverbal creativity is not as evenly divided among the two nonverbal tasks. The range of differences between the four grade level means is the greatest for the Circles tasks, although grade levels 6 and 8 had approximately equal means. The Incomplete Figures task shows generally smaller differences but a consistent increase with age.

The final graph in this series, Figure 25, depicts the mean scores for both boys and girls on the four tasks.

Figure 25

Scores on creativity tasks by sex




Despite the fact that the main effect of sex showed significant differences in favor of girls, boys achieved a higher mean performance score on both the Circles and Uses tasks. However, their superiority on these tasks was of less magnitude than the superiority achieved by the girls on the other two tasks. The graph tends to maximize the differences in favor of the boys. This is a function of rounding off the scores. On both Circles and Uses the differences were actually less then 1.0, while for Figure Completion and Product Improvement the difference in favor of the girls was somewhat over the 2.0 and 3.0, respectively. Of primary significance, however, is the fact that Figure 25 supports the assumption that none of the tasks are such as to favor one sex over the other. For example, if the comparative scores by sex had been roughly equal on three of the tasks and then greatly differentiated in favor of one sex on the fourth task, it would have been necessary to assume that the latter task was sex-biased.

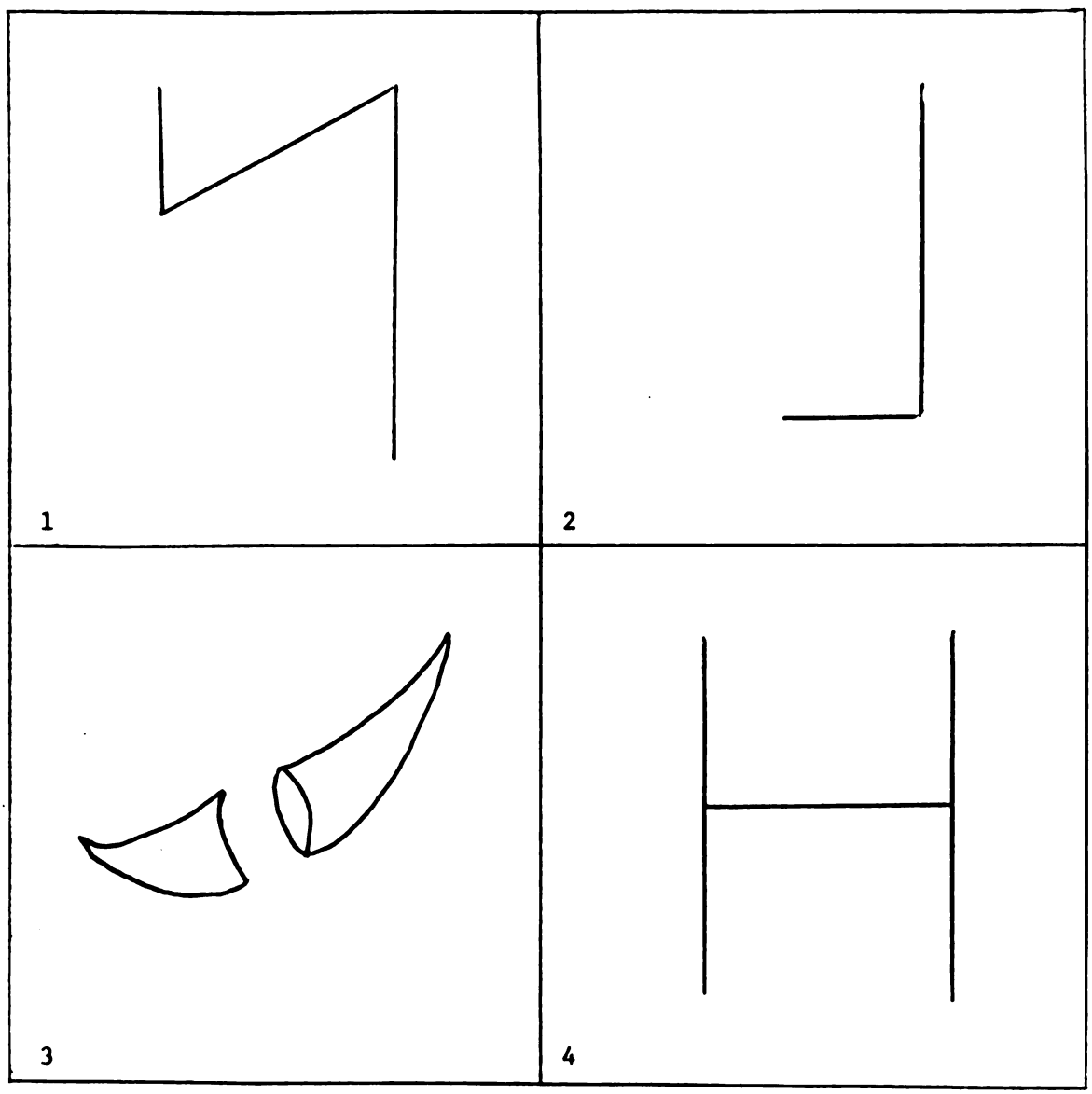
## Appendix D

Minnesota Test of Creative Thinking, Abbreviated Form VII

Appendix D contains copies of the four tasks that collectively comprise the Abbreviated Form VII, Minnesota Test of Creativity. They are: Incomplete Figures, Circles, Product Improvement and Unusual Uses of Tin Cans.

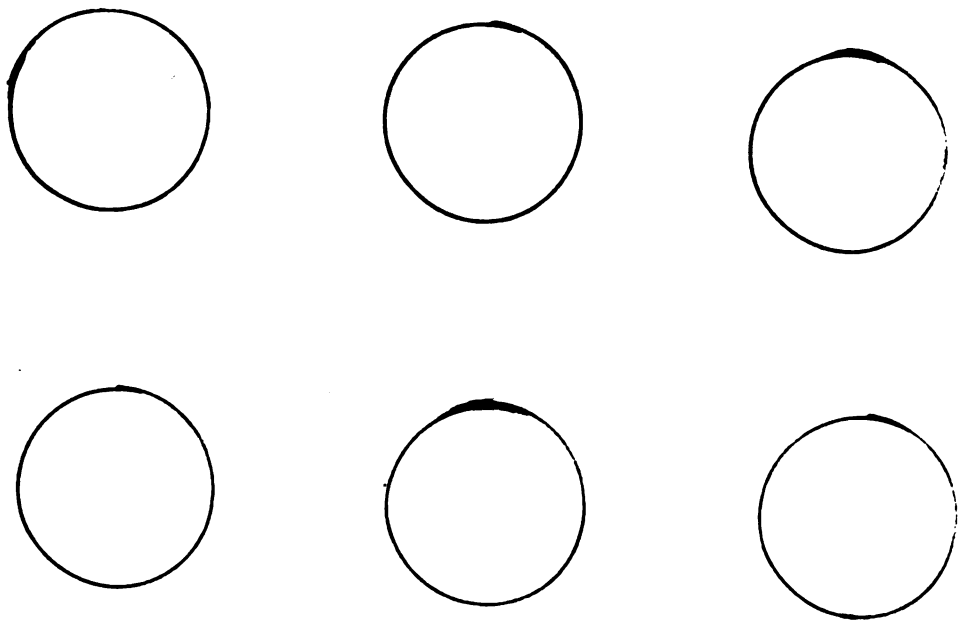
# Incomplete Figures

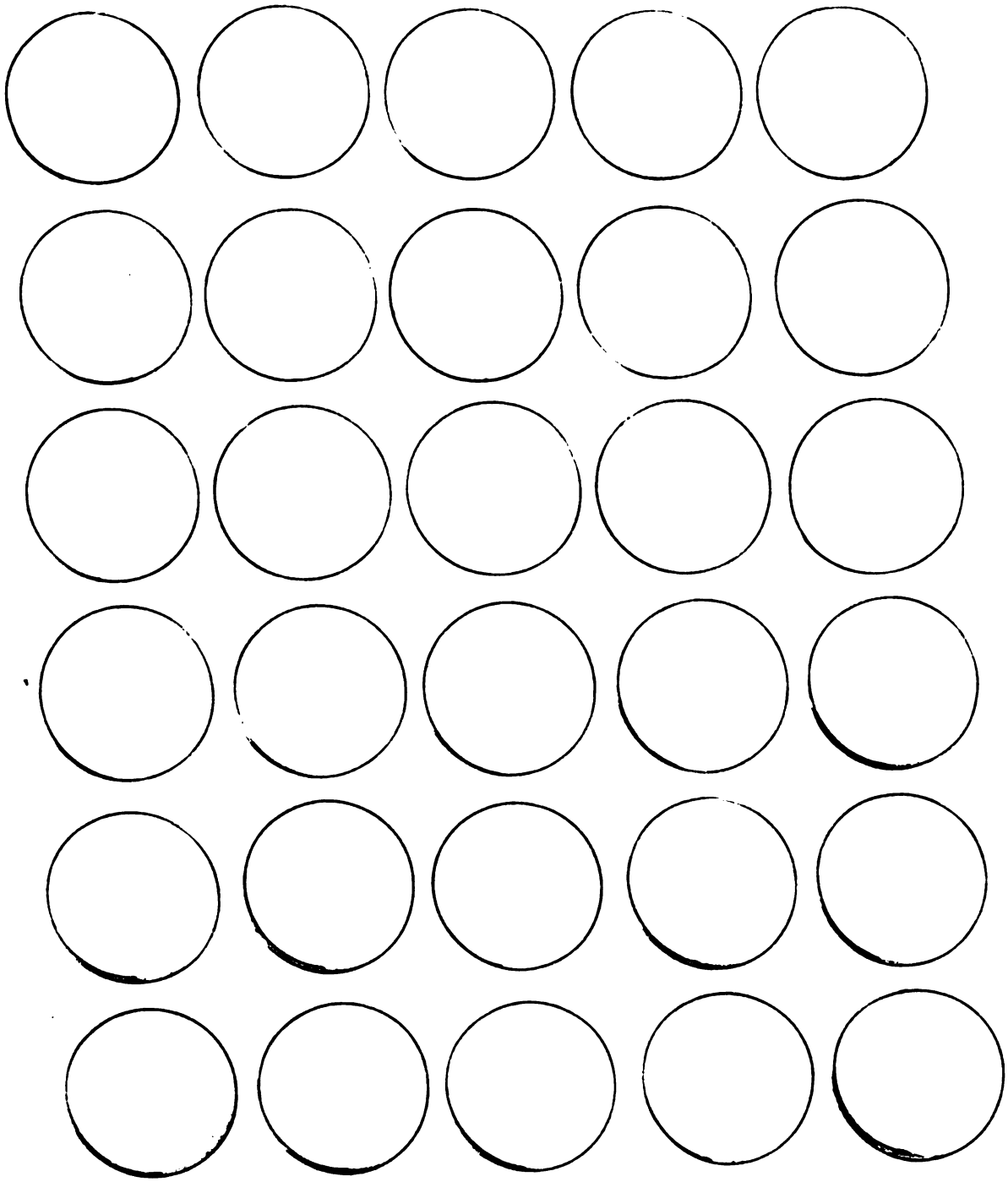
By adding lines to the figures on this and the next page, you can sketch some interesting objects or pictures. Again try to think of some picture or object that no one else will think of. Try to make it tell as complete and as interesting a story as you can by adding to and building up your first idea. Make up a title for each of your drawings and write at the bottom on each block next to the number of the figure.



# Circles

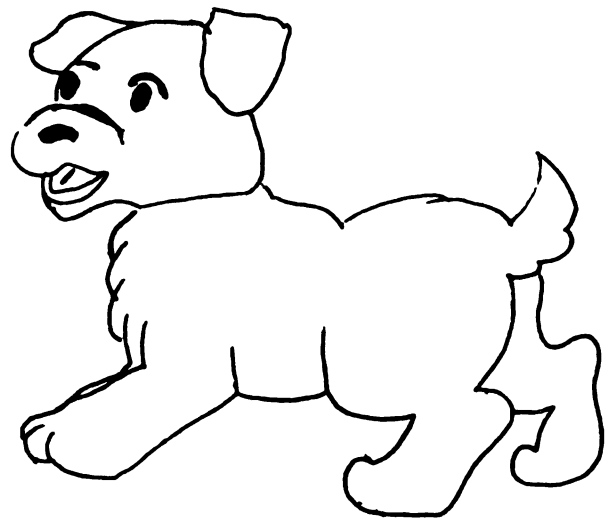
In ten minutes see how many objects or pictures you can make from the circles below and on the next page. The circles should be the main part of whatever you make. With pencil or crayon add lines to the circles to complete your picture. You can place marks inside the circles, outside the circles, or both inside and outside the circles -- wherever you want to in order to make your picture. Try to think of things that no one else will think of. Make as many different pictures or objects as you can and put as many ideas as you can in each one. Make them tell as complete and as interesting a story as you can. Add names or titles below the objects.





# Product Improvement

Try to think of the cleverest, most interesting and most unusual ways you can for changing this toy dog so that boys and girls will have more fun playing with it. Don't worry about how much it would cost -- just so it would make it more fun to play with. In the spaces on this page and on the next one, list the cleverest, most interesting and unusual ways you can think of for improving this toy dog.



# Unusual Uses of Tin Cans

Most people throw their empty tin cans away, but they have thousands of interesting and unusual uses. In the space below and on the next page, list as many of these interesting and unusual uses as you can think of. Do not limit yourself to any one size of can. You may use as many cans as you like. Do not limit yourself to the uses you have seen or heard about; think about as many possible new uses as you can.