THE USE OF THE WISC WITH EMOTIONALLY DISTURBED ADOLESCENTS

A Thesis

Presented to

The Department of Psychology

Kansas State Teachers College

In Partial Fulfillment

of the Requirements for the Degree

Master of Science

by

Steven J. Glicksman

December 1973

Thesis
1973
G

Approved for the Major Department

Approved for the Graduate Council

3150772

TABLE OF CONTENTS

Chapter Page

LIST OF TABLES

Table Page

Chapter 1

INTRODUCTION

The use of standardized tests has become one of the means by which a psychologist measures and evaluates the behavioral characteristics of an individual. These tests have been widely accepted by many psychologists as being valuable instruments for use with adults and with children. The study of psychometrics has shown that any test which is devised will not be a perfect measure, because human behavior is not completely understood. The standardization procedure attempts to eliminate as much chance error in the test as possible, by controlling the subjects upon whom the results are based. After the standardization procedure is completed, and the results compiled, the validity and the reliability of the test is found, in order to determine how much chance error there is. Then the psychologist will use this test on individuals whom he feels are most appropriately suited for it. Most standardized tests are best suited for those who fall in the middle range of the standardization sample. So that in an intelligence test, the scores of those who fall within the average range of intelligence are most reliable.

In our society, certain behavioral standards are set which our members are expected to abide by. Unfortunately, some individuals are unable to live by these standards because of a biological,

psychological and/or social limitation which is inherent. When administering a standardized test to any of these individuals, the psychologist must understand the reliability and the standardization sample it was based on, so that he can interpret the results and utilize them with regard to his subjects' limitations. It was found that when a test-retest method of estimating reliability of an intelligence test was done on a group of children with learning disabilities, the results were not consistent.[1] So that when a psychologist gives an intelligence test to a learning disabled child, he should interpret the results knowing that it is difficult, if not impossible, to find an accurate estimate of this child's intelligence.

THE PROBLEM

The study of emotionally disturbed children is a relatively new field. Before the twentieth century, a child who showed symptoms of what we now call an emotional disturbance would have his head drilled to let the evil spirits out; or he would be put in an institution where he would be treated inhumanely. It was not until the 1930's that it was recognized that these children have different needs than average children. Today, the techniques of working with these children and their behavioral characteristics are still vague, but extensive studies are now being conducted.

---

[1]James C. Coleman, "Stability of Intelligence Test Scores in Learning Disorders," Journal of Clinical Psychology, 19 (July, 1963), pp. 295-298.

Disturbed children can grow up in any social class and in any kind of family setting. Their behavior is often erratic and unpredictable, manifested by passive and/or aggressive reactions. These reactions vary from child to child and sometimes within one child. An emotional disturbance can create more problems to the child when he begins to learn through formal education. In the classroom, the disturbed child may show symptoms of his behavior which will prevent him from making adequate gains in his academic work. Therefore, the early diagnosis and treatment of these children is essential.

The Wechsler Intelligence Scale for Children (WISC)[2] is a standardized intelligence test which was developed by Wechsler in 1949. This test is considered by many psychologists to be one of the best individually administered intelligence tests.[3] The WISC consists of two scales: the Verbal and the Performance with each scale having five subtests and one supplementary test. The two scales are then combined into a Full Scale with an intelligence quotient (IQ) given for each scale and for the Full Scale. The test is analyzed with regard to the amount of scatter between the subtests and between the scales. The standardization of the WISC[4] was based on 2,200 white American children. One hundred

---

[2]Oscar Buros (ed.), "David Friedes," Volume 1, The Seventh Mental Measurements Yearbook (New Jersey: The Grayphon Press, 1972), p. 432.

[3]David Wechsler, Manual for the WISC (New York: The Psychological Corporation, 1949).

[4]Ibid.

boys and one hundred girls at each age from five to fifteen years old were chosen for the sample. Selection of the children was based on: (1) rural-urban residence; (2) father's occupation; and (3) geographic area. Based on the 1940 census and by the laws of stratified sampling, this group constituted an adequate sample of white children. The reliability for each subtest and for each scale was computed by the split-half technique. The results show that the reliability of the three scales denote a very high relationship, while the reliability of most subtests denote a moderate relationship. The clinician who interprets the results, must do so by understanding the implications of the reliability coefficients.

It is believed that because of the erratic and unpredictable behavior which is exhibited by those children who have been diagnosed as emotionally disturbed, the reliability of the use of the WISC on them is in question.

Statement of the Problem

Is there a significant difference in test-retest IQ scores obtained with the WISC with institutionalized emotionally disturbed adolescents and with normal adolescents?

Statement of the Hypothesis

There is no significant difference in test-retest IQ scores obtained with the WISC with institutionalized emotionally disturbed adolescents and with normal adolescents.

Purpose of the Study

The behavior of emotionally disturbed adolescents is so unpredictable that when a test is administered to them, their erratic behavior might affect the results. This study attempted to test the reliability of the WISC with emotionally disturbed adolescents utilizing a test-retest design. It was believed that if a significant difference was found in retest scores, the psychologist who tests these adolescents will have to interpret the test results with regard to the objective findings of this study.

Significance of the Study

The study of emotionally disturbed children is a relatively new field. The amount of research dealing directly with these children is minimal and this study attempted to add some very significant information. The WISC is a valuable measure for clinical and diagnostic purposes since it claims to evaluate specific abilities and pinpoint specific disabilities. It is believed that the results of this study can be used to aid the psychologist in his interpretation of the WISC scores of emotionally disturbed children.

DEFINITION OF TERMS

In order to help the reader understand what is meant by many of the terms used in this study, the following terms were defined.

Emotionally disturbed adolescents. Adolescents with emotional problems severe enough to prevent them from making the necessary adjustments for effective functioning in the culture.[5]

Standardized test. This is a test for which the validity and the reliability have been estimated.[6]

Reliability. "Refers to the extent to which a test is internally consistent. It also refers to how dependable the test is for predictive purposes." This term is usually defined in terms of a correlation coefficient. It is found by correlating two sets of measures.[7]

Stratified sampling.

> To classify the population with respect to the variable one wishes to control, and then to draw subjects from these classes so that they are represented in the same proportions in which they occur in the population.[8]

Population. The total group of people for which the test is to be used (e.g. all children from five to fifteen years old).

Split-half technique. A method of estimating reliability in which the items in a whole test are divided into two halves

---

[5]Charles W. Telford and James W. Saurey, The Exceptional Individual (New York: Holt, Reinhart and Winston, 1972), p. 421.

[6]Frank Freeman, Theory and Practice of Psychological Testing (third edition) (New York: Holt, Reinhart and Winston, 1962), p. 63.

[7]Ibid., pp. 66-68.

[8]Barry F. Anderson, The Psychology Experiment (California: Brooks/Cole Publishing Company, 1971), p. 48.

with each half treated as though it represented a separate form.[9] A correlation coefficient is then found from the results of each half.

Normal adolescents. Adolescents who are not in an institution or in any type of special education class.

Intelligence scale. ". . . is an assembled battery of tests; the intelligence rating obtained from them is a numerical expression of their combined contribution."[10]

Learning disability. ". . . refers to a marked discrepancy between the child's apparent potential and his performance level as he is engaged in essential learning processes."[11]

Mental retardation. ". . . refers to subaverage general intellectual functioning which originates during the developmental period and is associated with impairment in adaptive behavior."[12]

Exceptional children. Children who deviate from the norm in physical, mental, emotional, or social characteristics to such a degree that they require special social and educational services in order to develop their maximum capacity.[13]

---

[9]Freeman, op. cit., p. 69.

[10]David Wechsler, The Measurement and Appraisal of Adult Intelligence (Baltimore: The Williams and Wilkins Company, 1958), p. 16.

[11]Telford and Saurey, op. cit., p. 280.

[12]Ibid., p. 182.

[13]Ibid., p. 14.

LIMITATIONS OF THE STUDY

The population of this study was considered to be representative of the group described as emotionally disturbed, the the size of the sample is not totally accepted as being adequate. The larger the sample size, the more probable that it will yield results which are more representative of the population. The reader should keep in mind the sample size when interpreting the results.

Chapter 2

REVIEW OF RELATED LITERATURE

The use of the WISC with normal children has been found to be a reliable method of estimating these children's intellectual abilities.[1] The WISC is rarely given to the normal child. He is usually given a group intelligence test which is easier to administer and less time consuming. If a normal child shows some deviation on the group test, he will then be given the WISC or some other individual test of intelligence to confirm or reject the findings. The WISC, then, is generally administered to the exceptional child or to any child suspected of being exceptional. Littell[2] stated in 1962 that there have been no studies done dealing with the emotionally disturbed child. The bulk of research in this area has been done with the mentally retarded child, who is most often given this test.

THE RELIABILITY OF THE WISC WITH THE MENTALLY RETARDED

The use of intelligence tests for determining whether a child is mentally retarded is the main criterion for placing him

---

[1]Dale O. Irwin, "Reliability of the WISC," Journal of Educational Measurement, 3 (Winter, 1966), pp. 287-292.

[2]William Littell, "The WISC: Review of a Decade of Research," Psychology Bulletin, 57 (1960), pp. 132-156.

in a special class. It is therefore necessary that the reliability of the particular test used should be high, in order to make no mistake in placement. Children who are already in special classes for the mentally retarded are often retested to determine whether they should remain in these classes. Many studies have been done using a test-retest design to estimate the reliability of the WISC with mentally retarded children. Four of the studies (Whatley and Plant,[3] Walker and Gross,[4] Rosen et al,[5] Freidman[6]) have found no significant difference in test-retest scores. In these studies the interval time between tests was from two months (Walker and Gross) to six years (Rosen et al). Psychometricians believe that the optimum time for retesting an individual is between one and two weeks. When the interval time is more than two weeks, the following conditions might influence the reliability: (1) true difference in the subject (growth, practice); (2) skill in taking tests; (3) chance acquisition of knowledge; (4) coaching; (5) personal characteristics of the subject; (6) physical conditions

---

[3]Ruth G. Whatley and Walter T. Plant, "The Stability of WISC IQ's for Selected Children," Journal of Psychology, 44 (1957), pp. 165-167.

[4]Kenneth P. Walker and Fredrick L. Gross, "IQ Stability Among EMR Children," Training School Bulletin, 66, (4) February, 1970), pp. 181-187.

[5]Marvin Rosen, Linda Stallings, Lucretta Floor, and Myra Nowakiwska, "Reliability and Stability of Wechsler IQ Scores for Institutional Mental Subnormals," American Journal of Mental Deficiency, 73 (September, 1968), pp. 218-225.

[6]Ronald Freidman, "The Reliability of the WISC in a Group of Mentally Retarded Boys," Journal of Clinical Psychology, 26, (2) (April, 1970), pp. 181-182.

of the testing center; (7) chance factors, and (8) guessing.[7] It is, therefore, believed that although the results yielded no significant difference in test-retest scores, the interval times might have affected the results and invalidated the studies. The fifth condition warrants some consideration, because the behavior of the mentally retarded child is considerably different from the normal child.

> Although intellectual handicap and emotional maladjustment are clearly not related to each other in any simple fashion, the incidence of some degree of emotional disturbance is probably a great deal higher in retarded children than in children of average or superior intellect.[8]

Some of the causes for this high degree of emotional disturbance among the mentally retarded are: low frustration tolerance, poor home environment, disturbances in impulse control caused by brain damage, and abnormal psychological and physical development. When the mentally retarded child is considered also emotionally disturbed, the above studies seem to support the hypothesis of this study.

THE RELIABILITY OF THE WISC WITH LEARNING DISABILITY CHILDREN

The use of the WISC in diagnosing learning disabilities is a commonly used measure. The subtests of the WISC are related to many of the abilities taught in the schools and can give the

---

[7]Frank Freeman, Theory and Practice of Psychological Testing (third edition) (New York: Holt, Reinhart and Winston, 1962), p. 69.

[8]Herbert B. Robinson and Nancy M. Robinson, The Mentally Retarded Child (New York: McGraw-Hill Book Company, 1965), p. 223.

clinician some idea as to the areas of the child's disabilities. Children with learning disorders often have emotional problems because of their frustration in the classrooms. When a child is unable to read or do arithmetic because of an inherent disability, he is more apt to have a behavior problem because of his consistent failures. In a study done by Coleman,[9] he stated, "if the WISC does not yield a highly reliable estimate of a child's intellectual level, then serious sources of error may be introduced into both diagnosis and treatment of the child." His study was conducted with twenty-four subjects from the Psychology Clinic School for the Treatment of Learning Disorders. A test-retest procedure was used for finding the correlation between the two tests. The author concluded that the Full Scale IQ and the Performance IQ were not satisfactory for clinical assessment on an individual level. The Verbal IQ was below satisfactory and apparently resulted from the variance associated with the subject's characteristics. The mean test-retest interval was fifteen months, which is too long a period to accept his findings with confidence, but the correlation coefficients were high enough (FS = .77, PS = .81, and VS = .62), that the results cannot be overlooked.

---

[9]James C. Coleman, "Stability of Intelligence Test Scores in Learning Disorders," Journal of Clinical Psychology, 19 (July, 1963), pp. 295-298.

INTELLECTUAL ABILITY AND PERFORMANCE OF HYPERACTIVE CHILDREN

Many children who have been diagnosed as emotionally disturbed have a hyperkinetic disorder. This disorder "is a symptom complex of chronic, sustained, severe hyperactivity, marked distractability to extraneous stimuli, very short attention span, irritability and hyperexcitability."[10] One of the things known about this disorder is that it is sometimes caused by brain damage, but it is often difficult, if not impossible, to discern whether it is the damage to the brain which is causing this disorder. The brain damage is usually too small to be detected with a neurological examination and prevents an accurate diagnosis. When a child who is hyperactive is given the WISC, he often exhibits characteristics of his disorder which cause the results to be invalid. In a study conducted by Palkes and Stewart[11] in which thirty-two children from the St. Louis Children's Hospital Psychiatric Clinic having a hyperkinetic disorder were given the WISC, it was found that they were at a disadvantage in school and at home, not simply because of their behavior, but also because of an intellectual handicap. The level of intelligence, school achievement and perceptual motor performance was studied in the hyperactive children and compared with a control group matched for age, sex, race, grade

---

[10]Gerald Solomon, "Guidelines on the Use and Medical Effects of Psychostimulant Drugs in Therapy," Journal of Learning Disabilities, 35 (November, 1971), pp. 6-11.

[11]Helen Palkes and Mark Stewart, "Intellectual Ability and Performance of Hyperactive Children," American Journal of Orthopsychology, 42 (January, 1972), pp. 35-39.

and socioeconomic class. For each of the WISC scores, the mean of the control group was significantly higher than that for the hyperactive group. When the WISC is given to the hyperactive child, extreme care must be taken in the interpretation because his behavior might have affected the results.

THE STABILITY OF THE WISC IN A PSYCHIATRIC GROUP

The testing situation requires of the child many unfamiliar experiences. He is first asked into a strange room which is very bland in appearance. He is met by an examiner, who obviously has some authority and thereby is threatening to the child. They speak briefly about the test, the purpose and what is expected of the child. Then the child is asked many questions and is asked to play some games. Ideally, the examiner manipulates the situation with such precision that the child is put at ease and cooperates to his fullest capacity. Many times the examiner is not able to manipulate the situation so that the child is put at ease and cooperates. The emotionally disturbed child is usually so immature and insecure with himself, that when put into the testing situation he will show symptoms of his disorder. Test behavior is one of the things the psychologist considers when he interprets a test. When a child shows bizarre and/or erratic behavior in the testing situation, the psychologist must realize that the score obtained by his subject is representative of the child's abilities at the present time and not an accurate estimate at all times.

In a test-retest study of the WISC by Turner, Mathew and Rachman[12] involving thirty-four children from the In and Out Patient Clinics at the Children's Department of the Maudsley State Hospital (Great Britain), it was found that the means of the Full Scale and the Performance Scale tended to increase on retest, while the Verbal Scale tends to be insensitive to practice effects. The retest interval was six months, which, as was discussed earlier, might have affected the results. The children were classified into four diagnostic categories: (1) neurotic personality disorders, (2) anti-social conduct disorders, (3) habit disorders, and (4) unclassifiable. The above disorders are often considered to be caused by an emotional disturbance. All three scales were found to be of moderately high reliability. There was no significant difference found between the retest scores of the Verbal Scale, while there was a highly significant difference found for the Performance Scale, and a significant difference found for the Full Scale. The results of the above study tend to lean toward disagreement with the hypothesis of this study.

The use of the WISC with exceptional children is well accepted. The above studies seem to indicate that there is some discrepancy in the reliability of the Performance Scale and the Full Scale when used with certain groups of exceptional children.

---

[12]R. K. Turner, A. Mathews and S. Rachman, "The Stability of the WISC in a Psychiatric Group," British Journal of Educational Psychology, 37 (June, 1967), pp. 194-200.

Chapter 3

METHODS AND PROCEDURES

The methods and procedures by which this study was conducted will be discussed in order for another researcher to replicate the study. In order to determine if there was a significant difference between test-retest IQ scores obtained with the WISC between normal and disturbed adolescents, many variables were considered. The following discussion will help the reader to understand how the subjects were chosen, the methods by which they were tested and the ways the data were collected and analyzed.

POPULATION AND SAMPLING

Osawatomie State Hospital is a psychiatric hospital located one mile northeast of Osawatomie, Kansas, in the northeast section of the state. The hospital is managed by the Division of Institutional Management, under the State Board of Social Welfare, which is a part-time, policy making board of three members appointed by the Governor.

> The overall program and long-term goals of the hospital are to develop a modern accredited psychiatric facility which will meet American Psychiatric Association standards; to focus on care of the patient as an individual; to see that he is competently evaluated, diagnosed and treated; to make sure that he is adequately fed and comfortably housed in safe, sanitary and pleasant surroundings; and to work with the

> community toward the patient's total rehabilitation as a productive citizen.[1]

The patient population is urban, as well as rural, since the hospital serves twenty-three counties, including Johnson and Wyandotte counties (Kansas City and its surburbs). Patients under the age of eighteen years old are treated in a separate Adolescent Unit, which was established in March, 1964. Its program includes: individual and group psychotherapy, family therapy, biochemical therapy, academic classes, recreation and creative and work experiences with emphasis on teaching social competency, behavior skills and techniques.

Twenty adolescents were chosen for the study from the Adolescent Unit. The only criterion for being included in the study was that each adolescent was younger than fifteen years - eleven months (ceiling age for the WISC). At the time that the study was being conducted, there were, on the average, fifty adolescents on the unit most of whom were older than fifteen years - eleven months. These adolescents were admitted to the hospital for various psychological disorders. The psychiatric diagnosis of the subjects who were chosen for the study may be grouped according to the Diagnostic and Statistical Manual of Mental Disorders, second edition,[2] as follows: Personality Disorder (6);

---

[1]Dorothy C. Bishop, More Than One Hundred Years of Concern for the Mentally Ill, Public Information Office, Osawatomie State Hospital.

[2]Diagnostic and Statistical Manual of Mental Disorders (Washington, D.C.: American Psychiatric Association, second edition, 1968), pp. 88-102.

Psychosis (3); Transient Situational Disturbance (7); Behavior Disturbances of Childhood and Adolescence (4). There were nine boys and eleven girls in this group, whose mean age was fourteen years - eleven months; the range being thirteen years - six months to fifteen years - eleven months. The mean grade that these disturbed adolescents finished was 8.55; the range being seventh to tenth grade.

Twenty normal subjects were selected from Osawatomie, Kansas, and Emporia, Kansas. The adolescents were selected and matched with the disturbed group as to sex and age to serve as a control group. They were contacted with the use of lists which were provided by the local high schools. The criteria by which each normal adolescent was considered for the study was that he was younger than fifteen years - eleven months, that he had never been institutionalized for psychiatric treatment, or that he had never been in any kind of special education class. There were nine boys and eleven girls in this group, whose mean age was fourteen years - ten months; the range was from thirteen years - zero months to fifteen years - eleven months. The mean grade finished by these normal adolescents was 8.65, the range was from seventh to tenth grade.

MATERIALS AND INSTRUMENTATION

The WISC was administered to all forty adolescents twice. In order to eliminate examiner bias, two qualified examiners gave one test to each of the subjects. Each test was then scored by the examiner who administered the test. All of the subjects were

asked if they would mind taking an intelligence test for the purpose of helping the examiners with college course work. The adolescents did not know that they would be retested at a later date.

The WISC contains ten subtests and two supplementary tests; but in order to cut down the time involved, a short form of the test was administered to each subject. This particular short form was devised for use with emotionally disturbed children by Enburg, Rowley and Stone.[3] The authors stated the purposes of their study as being:

> . . . to devise short forms of the WISC such that (a) administration time could be lessened without appreciably reducing the reliability, and (b) previous sources of possible bias from preselection of subtest or utilization of Wechslers' original sample would be eliminated.

The results showed that there were four possible combinations of five subtests each (three Verbal and two Performance), which were suitable for use with emotionally disturbed children. The correlation between the five subtests and the sum of the Full Scale IQ for all four combinations was .96. The combination of five subtests which were chosen for this study was: Information, Comprehension, Arithmetic, Picture Completion, and Block Design. The reason that this particular combination was chosen was because these subtests were most often included in the other three combinations.

---

[3]Richard Enburg, Vinton N. Rowley and Beth Stone, "Short Forms of the WISC for Use with Emotionally Disturbed Children," Journal of Clinical Psychology, 17 (July, 1961), pp. 280-284.

A description of what these five subtests measure is as follows:[4]

Information. This subtest measures how much general information the subject has abstracted from his surrounding environment. It calls into operation remote memory, ability to comprehend, capacity for associative thinking as well as the interests and reading background of the subject.

Comprehension. This subtest measures the level of the subject's ability to use practical judgment in everyday social actions, the extent to which social acculturation has taken place, and the extent to which a maturing conscience or moral sense has developed.

Arithmetic. This subtest measures the ability to utilize abstract concepts of number and numerical operations which require complex thought patterns.

Picture completion. This subtest calls for visual identification of familiar objects, forms, and living things and the further capacity to identify and isolate essential from non-essential characteristics. Attention and concentration are important elements in the test.

---

[4]Alan J. Glasser and Irla Lee Zimmerman, Clinical Interpretation of the WISC (New York and London: Grune and Stratton, 1967), pp. 36-95.

Block design. This subtest measures perception, analysis, synthesis and reproduction of abstract designs. Visual-motor coordination is also measured.

DESIGN OF THE STUDY

The test-retest design (using the same measure) was used in this study to determine if there is a significant difference between the two groups (normal and disturbed) on retest with the WISC. This design is a commonly used measure for estimating reliability of standardized tests. There are many disadvantages in using this method, but this study attempted to control as many of them as possible.

One criticism is that the same test is given twice, and the subject may do better on his second try since he was familiar with the test. It was found[5] that scores obtained with the WISC tend to increase significantly on retest at all ages, and that the Performance subtests are much more susceptible to short-term practice effects than Verbal subtests. With the use of a control group, it seems insignificant whether practice effects have had any bearing upon the results because both groups had the same amount of practice.

One of the main criticisms of this method is that if the retest is given later than two weeks, the reliability of the second test is questionable. This study attempted to control this time

---

[5]Mohammed Y. Quereshi, "Practice Effects on the WISC Subtests and IQ Estimates," Journal of Clinical Psychology, 24 (January, 1968), pp. 79-85.

factor by retesting most subjects within the recommended interval of from one to two weeks. The mean time (in days) for the normal group was six, whereas, the mean time for the disturbed group was 11.6. The range of the normal group was from seven to sixteen days, and the range for the disturbed group was from two to twenty-four days.

DATA COLLECTION

When each adolescent showed up for testing, the examiner immediately tried to establish rapport. Each subject was told that he would be given an intelligence test and that the results would not be used for clinical or diagnostic purposes. It was explained that the examiner was a graduate student in psychology and that it was necessary to give a certain number of tests as practice for a course he was taking. The subject was also told that he was not expected to know all the answers because no one knows all the answers to everything; but he was expected to do the best he could. Standardized procedures were followed by both examiners in administering the tests.

DATA ANALYSIS

The data in this study were analyzed in order to determine if there was a significant difference in test-retest IQ scores obtained with the WISC with institutionalized emotionally disturbed adolescents and with normal adolescents.

t-Test

The first step in the treatment of the data was to determine the difference between the pre- and posttests for each group on all three scales (Verbal, Performance and Full Scale). It was also necessary to determine the difference between the pre- and posttests for each group on all five scales (Information, Comprehension, Arithmetic, Picture Completion and Block Design). To make this determination, the standard deviation, the mean, the correlation, and the mean difference for each group on all three scales, and on all five subtests on the pre- and posttests had to be calculated. In order to reject or accept the null hypothesis at the .01 level of significance, it was necessary to determine also the degrees of freedom to be used. The degrees of freedom for this t-test was calculated by N - 1. To show these differences effectively, a t-table was set up showing the pertinent values that were determined. The formula that was used to determine the t-values was as follows:

$$t = \frac{\bar{X}_{pre} - \bar{X}_{post}}{\sqrt{\frac{S^2_{pre}}{n^2_{pre}} + \frac{S^2_{post}}{n^2_{post}} - 2r \frac{S_{pre}}{n_{pre}} \frac{S_{post}}{n_{post}}}}$$

where:

$\bar{X}_{pre}$ = mean value of the pretest

$\bar{X}_{post}$ = mean value of the posttest

S^2_{pre} = variance of the pretest

S_{pre} = standard deviation of the pretest

S^2_{post} = variance of the posttest

S_{post} = standard deviation of the posttest

n_{pre} = number on the pretest

n_{post} = number on the posttest

r = Pearson Product Moment Coefficient of Correlation

Analysis of Covariance

The second test utilized in the statistical analysis of the data was that of analysis of covariance. In its most basic form, one might think of analysis of covariance first by determining the magnitude and the direction of the relationship between the control variable and the criterion variable. The adjusted total sum of squares was calculated from:

$$SS'_{ty} = SS_{ty} - \frac{(SP_t)^2}{SS_{wx}}$$

where:

SS'_{ty} = adjusted total sum of squares

SS_{ty} = total sum of squares for the Y-scores

SP_t = total sum of products for the X- and Y-scores

SS_{wx} = sum of squares for within groups of the X-scores

Similarly, the adjusted sum of squares for within groups was determined from:

$$SS'_{wy} = SS_{wy} - \frac{(SP_w)^2}{SS_{wx}}$$

where:

SS'_{wy} = adjusted sum of squares for within groups

SS_{wy} = sum of squares for within groups for the Y-scores

SP_w = sum of products for within groups

Finally, the adjusted sum of squares between groups was calculated as a residual:

$$SS'_{by} = SS'_{ty} - SS'_{wy}$$

where:

SS'_{by} = adjusted sum of squares between groups

As in the simple analysis of variance, the number of degrees of freedom for the adjusted sum of squares between groups is (k - 1). However, one degree of freedom is lost by imposing the restriction that the deviations be computed from the common within groups regression line, and the number of degrees of freedom for the adjusted sum of squares is (N - k - 1), and for the adjusted total sum of squares is (N - 2). The adjusted mean squares between and within are obtained by dividing the sums of squares by their respective degrees of freedom. The test of the hypothesis of equal means is obtained from:

$$F = \frac{MS'_{by}}{MS'_{wy}} \quad \text{... with df} = (k - 1), (N - k - 1)^6$$

where:

MS'_{by} = adjusted mean square value obtained for between groups

MS'_{wy} = adjusted mean square value obtained for within groups

The adjusted mean formula for the Y-measures is found by:

$$Y' = Y - b_w (X - M_x)$$

---

[6]John T. Roscoe, Fundamental Research Statistics for Behavioral Sciences (New York: Holt, Reinhart, and Winston, Incorporated, 1969), p. 257.

where:

$\bar{Y}$ = original mean of the Y-scores

b_w = slope of the line (the rate Y changes with the change in X)

$\bar{X}$ = the original mean of the X-scores

M_X = overall mean of the X-scores

The value of b_w (the slope of the line) is calculated by the formula:

$$b_w = \frac{SP_w}{SS_{wx}}$$

where:

SP_w = sum of products for within groups (error)

SS_{wx} = sum of products for within groups (error) for X scores

Chapter 4

ANALYSIS OF DATA

The following statistical data will help the reader to understand the results of the study.

STATISTICAL ANALYSIS

A *t*-test was used to determine if there was a significant difference between the two populations who participated in this study on pre- and posttest IQ scores, obtained with the WISC. The *t*-test formula can be found on page 23 of Chapter 3.

Disturbed Group: Pre- and Post-Test Information Subtest

There were twenty adolescents in this group, who had a mean scale score of 6.95 and a standard deviation of 1.72 on the

Table 1

t-Test for Disturbed Group on Pretest and Posttest of the Information Subtest

Group	Number	Standard Deviation	Mean	Mean Difference	Degrees of Freedom	*r*	*t*
Pre	20	2.52	10.45	-.75	19	0.854	*-2.51
Post	20	2.42	11.20				

*Significant at the .01 level of significance.

pretest. On the posttest, the standard deviation was 2.01 and the mean scale score was 8.20.

Using nineteen degrees of freedom (df = N - 1), a *t*-value greater than or equal to ±2.861 ($t \geq \pm 2.861$) was needed to reject the null hypothesis at the .01 level of significance. With an obtained mean difference of -1.25, a *t*-value of -4.93 was calculated. Since the obtained value fell within the critical region, the null hypothesis was rejected. It was concluded that the disturbed group did significantly better on the posttest of the Information subtest.

A correlation of .827 (r = 0.827) was calculated between the pre- and posttests, with eighteen degrees of freedom. It was significant at the .01 level and is a moderately high degree of relationship.

Disturbed Group: Pre- and Post-Test Comprehension Subtest

In this group of twenty adolescents, a mean scale score of 6.60 was obtained, with a standard deviation of 2.35 on the

Table 2

t-Test for Disturbed Group on Pretest and Posttest of the Comprehension Subtest

Group	Number	Standard Deviation	Mean	Mean Difference	Degrees of Freedom	*r*	*t*
Pre	20	2.35	6.60	-.75	19	0.639	-1.72
Post	20	2.22	7.35				

pretest. On the posttest, the standard deviation was 2.22 and the mean scale score was 7.35.

A *t*-value of -1.72 was computed with a mean difference of -.75. Since the obtained value of -1.72 did not fall within the critical region, the null hypothesis was not rejected. It was concluded that the disturbed group did not do significantly better on the posttest of the Comprehension subtest.

Disturbed Group: Pre- and Post-Test Arithmetic Subtest

In this group of twenty adolescents, a mean scale score of 8.05 was obtained with a standard deviation of 3.11 on the pretest. On the posttest, the standard deviation was 2.56 and the mean scale score was 8.50.

Table 3

t-Test for Disturbed Group on Pretest and Posttest of the Arithmetic Subtest

Group	Number	Standard Deviation	Mean	Mean Difference	Degrees of Freedom	*r*	*t*
Pre	20	3.11	8.05	-.45	19	0.752	-.98
Post	20	2.56	8.50				

With an obtained mean difference of -.45, a *t*-value of -.98 was computed. Because the obtained value did not fall in the critical region, the null hypothesis was not rejected. It was concluded that the disturbed group did not do significantly better on the posttest of the Arithmetic subtest.

A correlation of .752 (r = 0.752) was calculated between the pre- and posttests. This correlation was significant at the .01 level and is a moderately high degree of relationship.

Disturbed Group: Pre- and Post-Test Picture Completion Subtest

A standard deviation of 3.00 and a mean scale score of 10.35 was computed on the Picture Completion subtest for this group of twenty adolescents. A mean scale score of 11.55 was calculated on the posttest and 3.46 was computed as the standard deviation.

Table 4

t-Test for Disturbed Group on Pretest and Post-Test of the Picture Completion Subtest

Group	Number	Standard Deviation	Mean	Mean Difference	Degrees of Freedom	*r*	*t*
Pre	20	3.00	10.35	-1.20	19	0.766	*-2.39
Post	20	3.46	11.55				

*Significant at the .05 level of significance.

A *t*-value of -2.39 was calculated with a mean difference of -1.20. Since the obtained value fell within the critical region, the null hypothesis was rejected. It was concluded that the disturbed group did significantly better on the posttest of the Picture Completion subtest.

A correlation of .766 (r = 0.766) was calculated between the pre- and posttests, which was significant at the .01 level and is considered a moderately high degree of relationship.

Disturbed Group: Pre- and Post-Test Block Design Subtest

There were twenty adolescents in this group, who had a mean scale score of 9.75 and a standard deviation of 3.43 on the pretest. On the posttest, the standard deviation was 4.01 and a mean scale score of 10.75 was calculated.

Table 5

t-Test for Disturbed Group on Pretest and Posttest of the Block Design Subtest

Group	Number	Standard Deviation	Mean	Mean Difference	Degrees of Freedom	*r*	*t*
Pre	20	3.43	9.75	-1.00	19	0.881	*-2.36
Post	20	4.01	10.75				

*Significant at the .05 level of significance.

With an obtained mean difference of -1.00, a *t*-value of -2.36 was computed. Because the obtained value of -2.36 fell within the critical region, the null hypothesis was rejected. It was concluded that the disturbed group did significantly better on the posttest of the Block Design subtest.

A correlation of .881 ($r = 0.881$) was calculated between the pre- and posttests. It was significant at the .01 level and would generally be interpreted as a moderately high degree of relationship.

Disturbed Group: Pre- and Posttest Verbal IQ

In this group of twenty adolescents, a mean Verbal IQ of 82.40 was obtained with a standard deviation of 9.76 on the pretest. On the posttest, the standard deviation was calculated to be 9.98 and a mean Verbal IQ of 87.60 was computed.

Table 6

t-Test for Disturbed Group on Pretest and Posttest Verbal IQ

Group	Number	Standard Deviation	Mean	Mean Difference	Degrees of Freedom	*r*	*t*
Pre	20	9.76	82.40	-5.20	19	0.702	*-2.97
Post	20	9.98	87.60				

*Significant at the .01 level of significance.

With an obtained mean difference of -5.20, a *t*-value of -2.97 was calculated. Since the obtained value of -2.97 fell within the critical region, the null hypothesis was rejected. It was concluded that the disturbed group did significantly better on the posttest of the Verbal IQ.

A correlation of .702 ($r = 0.702$) was established. It was significant at the .01 level and would generally be interpreted as a moderately high degree of relationship.

Disturbed Group: Pre- and Posttest Performance IQ

A standard deviation of 17.45 and a mean value of 101.20 was obtained on the Performance section of the test for this group of twenty adolescents. A mean value of 107.40 was calculated on the posttest and 23.79 was computed as the standard deviation.

Table 7

t-Test for Disturbed Group on Pretest and Posttest Performance IQ

Group	Number	Standard Deviation	Mean	Mean Difference	Degrees of Freedom	*r*	*t*
Pre	20	17.45	101.20	-6.20	19	0.816	-1.95
Post	20	23.79	107.40				

With an obtained mean difference of -6.20, a *t*-value of -1.95 was computed. Since the obtained value of -1.95 did not fall within the critical region, the null hypothesis was not rejected. It was concluded that the disturbed group did not do significantly better on the posttest of the Performance section of the test.

A correlation of .816 ($r = 0.816$) was calculated between the pre- and posttests, which was significant at the .01 level and is a moderately high degree of relationship.

Disturbed Group: Pre- and Posttest Full Scale IQ

In this group of twenty adolescents, a mean Full Scale IQ of 90.50 was obtained with a standard deviation of 12.74 on the pretest. On the posttest, the standard deviation was 15.09 with a mean Full Scale IQ of 96.60.

Table 8

t-Test for Disturbed Group on Pretest and Posttest Full Scale IQ

Group	Number	Standard Deviation	Mean	Mean Difference	Degrees of Freedom	*r*	*t*
Pre	20	12.74	90.50	-6.10	19	0.791	*-2.87
Post	20	15.09	96.60				

*Significant at the .01 level of significance.

With an obtained mean difference of -6.10, a *t*-value of -2.87 was calculated. Since the obtained value of -2.87 fell within the critical region, the null hypothesis was rejected. It was concluded that the disturbed group did significantly better on the posttest of the Full Scale IQ.

Normal Group: Pre- and Post-Test Information Subtest

In this group of twenty adolescents, a mean scale score of 10.45 was obtained with a standard deviation of 2.52 on the pretest. On the posttest, the standard deviation was 2.42 and the mean scale score was 11.20.

Table 9

t-Test for Normal Group on Pretest and Posttest of the Information Subtest

Group	Number	Standard Deviation	Mean	Mean Difference	Degrees of Freedom	*r*	*t*
Pre	20	2.52	10.45	-.75	19	0.854	*-2.51
Post	20	2.42	11.20				

*Significant at the .05 level of significance.

A *t*-value of -2.51 was computed with a mean difference of -.75. Because the obtained *t*-value of -2.51 fell within the critical region, the null hypothesis was rejected. It was concluded that the normal group did significantly better on the posttest of the Information subtest.

A correlation of .854 (r = 0.854) was established. It was significant at the .01 level and would generally be interpreted as a moderately high degree of relationship.

Normal Group: Pre- and Post-Test Comprehension Subtest

A standard deviation of 3.76 and a mean scale score of 9.65 was obtained on the Comprehension subtest for this group of twenty adolescents. A mean scale score of 10.90 was computed on the posttest and 3.63 was calculated as the standard deviation.

Table 10

t-Test for Normal Group on Pretest and Post-Test of the Comprehension Subtest

Group	Number	Standard Deviation	Mean	Mean Difference	Degrees of Freedom	*r*	*t*
Pre	20	3.76	9.65	-1.25	19	0.862	*-2.87
Post	20	3.63	10.90				

*Significant at the .01 level of significance.

With an obtained mean difference of -1.25, a *t*-value of -2.87 was computed. Since the obtained value of -2.87 fell within the critical region, the null hypothesis was rejected. It was concluded that the normal group did significantly better on the posttest of the Comprehension subtest.

A correlation of .862 (r = 0.862) was calculated between the pre- and posttests, which was significant at the .01 level and is a moderately high degree of correlation.

Normal Group: Pre- and Post-Test Arithmetic Subtest

In this group of twenty adolescents, a mean scale score of 11.20 was obtained with a standard deviation of 2.66 on the pretest. On the posttest, the standard deviation was computed to be 3.27 and a mean scale score of 12.10 was calculated.

With an obtained mean difference of -.90, a *t*-value of -2.41 was calculated. Since the obtained value of -2.41 fell within the critical region, the null hypothesis was rejected. It

Table 11

t-Test for the Normal Group on Pretest and Posttest of the Arithmetic Subtest

Group	Number	Standard Deviation	Mean	Mean Difference	Degrees of Freedom	*r*	*t*
Pre	20	2.66	11.20	-.90	19	0.861	*-2.41
Post	20	3.27	12.10				

*Significant at the .05 level of significance.

was concluded that the normal group did significantly better on the Arithmetic subtest upon retest.

A correlation of .861 (r = 0.861) was established. It was significant at the .01 level and would generally be interpreted as a moderately high degree of relationship.

Normal Group: Pre- and Posttest
Picture Completion Subtest

In this group of twenty adolescents, a mean scale score of 10.50 was obtained with a standard deviation of 2.54 on the pretest. On the posttest, the standard deviation was 2.37 and the mean scale score was 11.50.

With an obtained mean difference of -1.00, a *t*-value of -2.67 was calculated. Since the obtained value of -2.67 fell within the critical region, the null hypothesis was rejected. It was concluded that the normal group did significantly better on the posttest of the Picture Completion subtest.

Table 12

t-Test for the Normal Group on the Pretest and Posttest of the Picture Completion Subtest

Group	Number	Standard Deviation	Mean	Mean Difference	Degrees of Freedom	r	t
Pre	20	2.54	10.50	-1.00	19	0.770	*-2.67
Post	20	2.37	11.50				

*Significant at the .05 level of significance.

A correlation of .770 (r = 0.770) was established. It was significant at the .01 level and would generally be interpreted as a moderately high degree of relationship.

Normal Group: Pre- and Post-Test Block Design Subtest

A standard deviation of 3.28 and a mean scale score of 12.10 was obtained on the Block Design subtest for this group of

Table 13

t-Test for Normal Group on Pretest and Posttest of the Block Design Subtest

Group	Number	Standard Deviation	Mean	Mean Difference	Degrees of Freedom	r	t
Pre	20	3.28	12.10	-1.00	19	0.867	*-2.58
Post	20	3.41	13.10				

*Significant at the .05 level of significance.

twenty adolescents on the pretest. On the posttest, a mean scale score of 13.10 was computed and a standard deviation of 3.41 was calculated.

A *t*-value of -2.36 was computed with a mean difference of -1.00. Because the obtained value of -2.36 fell within the critical region, the null hypothesis was rejected. It was concluded that the normal group did significantly better on the posttest of the Block Design subtest.

A correlation of .881 ($r = 0.881$) was calculated between the pre- and posttests. It was significant at the .01 level and would generally be interpreted as a moderately high degree of relationship.

Normal Group: Pre- and Posttest Verbal IQ

A standard deviation of 12.69 and a mean value of 103.70 was obtained on the Verbal section of the test for this group of

Table 14

t-Test for Normal Group on Pretest and Posttest Verbal IQ

Group	Number	Standard Deviation	Mean	Mean Difference	Degrees of Freedom	*r*	*t*
Pre	20	12.69	103.20	-5.65	19	0.846	*-3.20
Post	20	14.39	108.90				

*Significant at the .01 level of significance.

twenty adolescents. A mean value of 108.90 was calculated on the posttest and the standard deviation was 14.39.

With an obtained mean difference of -5.65, a t-value of -3.20 was computed. Since the obtained value of -3.20 fell within the critical region, the null hypothesis was rejected. It was concluded that the normal group did significantly better on the posttest of the Verbal section of the test.

A correlation of .846 ($r = 0.846$) was established. It was significant at the .01 level and would generally be interpreted as a moderately high degree of relationship.

Normal Group: Pre- and Post-Test Performance IQ

A standard deviation of 17.06 and a mean value of 109.10 was obtained on the Performance section of the test for this group of twenty adolescents. A mean value of 115.90 was calculated on the posttest and the standard deviation was 16.53.

Table 15

t-Test for Normal Group on Pretest and Posttest Performance IQ

Group	Number	Standard Deviation	Mean	Mean Difference	Degrees of Freedom	r	t
Pre	20	17.06	109.10	-6.80	19	0.853	*-3.25
Post	20	16.53	115.90				

*Significant at the .01 level of significance.

With an obtained mean difference of -6.80, a t-value of -3.25 was computed. Since the obtained value fell within the critical region, the null hypothesis was rejected. It was concluded that the normal group did significantly better on the posttest of the Performance section of the test.

A correlation of .853 (r = 0.853) was established. It was significant at the .01 level and would be generally interpreted as a moderately high degree of relationship.

Normal Group: Pre- and Post-Test Full Scale IQ

In this group of twenty adolescents, a mean Full Scale IQ of 106.70 was obtained with a standard deviation of 15.11 on the pretest. On the posttest, the standard deviation was 15.86 and a mean Full Scale IQ of 113.25 was calculated.

Table 16

t-Test for Normal Group on Pretest and Posttest Full Scale IQ

Group	Number	Standard Deviation	Mean	Mean Difference	Degrees of Freedom	r	t
Pre	20	15.11	106.70	-6.55	19	0.910	*-4.33
Post	20	15.86	113.25				

*Significant at the .01 level of significance.

With an obtained mean difference of -6.55, a t-value of -4.33 was computed. Since the obtained value of -4.33 fell within the critical region, the null hypothesis was rejected. It was

concluded that the normal group did significantly better on the posttest of the Full Scale IQ.

A correlation of .910 (r = 0.910) was calculated between the pre- and posttests which was significant at the .01 level and is an extremely high correlation.

Analysis of Covariance

The second part of the statistical analysis in this study was that of analysis of covariance. In brief, analysis of covariance may be used when a relationship is being studied between a dependent variable and two or more groups representing an independent variable. This technique allows the researcher to statistically equate the independent variables with respect to one or more variables which are relevent to the dependent variable. The formula for Analysis of Covariance can be found on page 24 of Chapter 3.

Normal and Disturbed Group: Performance IQ

The sum of squares of the X-scores between groups was 631. The sum of squares for between groups for the Y-scores was 731. The sum of products for the X and Y-scores was 680.

Within the groups, the sum of squares for the X-scores was 11,917. It was computed that the sum of squares for the Y-scores was 18,784. The sum of products for the X and Y-scores was 11,595. The F-table was used at one and thirty-seven degrees of freedom at the .05 level of significance.

In analyzing the table, the adjusted sum of squares for the Y-scores between groups was -631, and the adjusted mean square

Table 17

Analysis of Covariance for Disturbed and Normal Groups on the Pretest and Posttest Performance IQ of the WISC

Source	df	SS_x	SP	SS_y	df'	SS'_y	MS'_y
Between	1	631	680	731	1	-631	-631
Within	38	11,917	11,595	18,784	37	7,502	203
Total	39	12,548	12,275	19,515	38	6,871	

for the Y-scores was -631. Within the groups, the adjusted sum of squares for the Y-scores was 7,502. The adjusted mean square for the Y-scores was 203.

The F-table was used at one and thirty-seven degrees of freedom. An F-value greater than or equal to ± 4.17 (F_1, 37 ± 4.17) at the .05 level of significance was needed to reject the null hypothesis. Since a value of -3.112 was obtained, this value was not considered significant. Therefore, it would be concluded that there was no significant difference between the adjusted means of the Performance IQ's of the normal and disturbed groups on retest.

Normal and Disturbed Groups:
Verbal IQ

Table 18 shows that the sum of squares for the X-scores between groups was 4,326. The sum of squares for the Y-scores for between groups was 4,516. It shows that the sum of the products for the X and Y-scores was 4,420.

The sum of squares of the X-scores for within groups was 5,128. Within the groups, the sum of squares of the Y-scores was 6,131. It was calculated that the sum of the products for the X and Y-scores was 4,457.

Table 18

Analysis of Covariance for Disturbed and Normal Groups on the Pretest and Posttest Verbal IQ of the WISC

Source	df	SS_x	SP	SS_y	df'	SS'_y	MS'_y
Between	1	4,326	4,420	4,516	1	-6,977	-6,977
Within	38	5,128	4,457	6,131	37	2,257	61
Total	39	9,454	8,877	10,647	38	-4,720	

The analysis of Table 18 shows the adjusted sum of squares to be -6,977, for between groups and the adjusted mean square to be -6,977. Within the groups, the adjusted sum of squares was 2,257, while the adjusted mean square was 61.

With an obtained F-value of -114.37, the null hypothesis was rejected. It was concluded that there was a significant difference between the adjusted means of the Verbal IQ's of the normal and disturbed groups.

Normal and Disturbed Groups:
Full Scale IQ

The sum of squares of the X-scores between the groups was 2,625. The sum of squares for between the groups for the Y-scores was 2,772. The sum of products for the X and Y-scores was 2,697.

Within the groups, the sum of squares for the X-scores was 7,811. It was computed that the sum of squares for the Y-scores was 9,587. The sum of products for the X and Y-scores was 7,407.

Table 19

Analysis of Covariance for Disturbed and Normal Groups on the Pretest and Posttest Full Scale IQ of the WISC

Source	df	SS_x	SP	SS_y	df'	SS'_y	MS'_y
Between	1	2,625	2,697	2,772	1	-3,274	-3,274
Within	38	7,811	7,407	9,587	37	2,563	69
Total	39	10,436	10,104	12,359	38	-711	

In analyzing the table, the adjusted sum of squares for the Y-scores between groups was -3,274 and the adjusted mean square for the Y-scores was -3,274. Within the groups, the adjusted sum of squares for the Y-scores was 2,563. The adjusted mean square for the Y-scores was 69.

Since a value of -47.27 was obtained, this value was considered significant. Therefore, it would be concluded that there was a significant difference between the adjusted means of the Full Scale IQ's of the normal and disturbed groups on retest.

Chapter 5

SUMMARY, CONCLUSIONS AND RECOMMENDATIONS

The use of the WISC with emotionally disturbed children is a respected instrument in diagnosis and remediation of these children's disturbances. It was believed that if there was a significant difference between test-retest scores obtained with the WISC with disturbed children, then the psychologist who is diagnosing and treating these children must be extremely careful in interpreting the results.

SUMMARY

The behavior of children who have been diagnosed as having an emotional disturbance is so inconsistent that it is questionable whether a standardized test yields information which is truly an accurate representation of these individuals. When a child is troubled with himself or with his environment, it would seem probable that his abilities to perform in situations calling for concentrated thought and manual dexterity are detered by this emotionality. It was, therefore, believed that if a standardized test was administered to a group of adolescents who have been institutionalized for an emotional disturbance and readministered to them within a two week interval, the results would show just how stable this test was.

The Wechsler Intelligence Test for Children (WISC), is a standardized intelligence test which is considered one of the best individually administered tests of its kind. The test contains three scales: Verbal and Performance, each scale having five subtests and two supplementary tests; and a third scale, the Full Scale, which is a combination of the other two scales. An intelligence quotient (IQ) is found for all three scales.

Twenty subjects from the Adolescent Unit at Osawatomie State Hospital and twenty presumed to be normal subjects (who served as a control group) from Osawatomie, Kansas, and Emporia, Kansas, were given a short form of the WISC by two qualified examiners. This particular short form was especially devised for use with emotionally disturbed children. All subjects were asked if they would mind taking an intelligence test to help the examiners with college course work. None of the adolescents were told that they would be retested at a later date.

There were nine boys and eleven girls in each group who took part in the study. The mean age of the disturbed group was fourteen years - eleven months, with a range from thirteen years - six months to fifteen years - eleven months. The mean age of the normal group was fourteen years - ten months, with a range from thirteen years - zero months to fifteen years - eleven months. The adolescents in the disturbed group had a mean grade completion of 8.55, while the normal group had a mean grade completion of 8.65. Both groups had a grade range of from seventh to tenth grade. The mean time (in days) for retest for the disturbed group was 11.60, with a range from two to twenty-four days. The mean time

for retest for the normal group was six, with a range from seven to sixteen days.

The first part in the statistical analysis of the data was the t-test. This test was used to determine if there was a significant difference between pre- and posttest scores for both groups. The t-test was used first with the five subtests given. It was concluded that the normal group did significantly better at the .05 level on four subtests (Information, Arithmetic, Picture Completion, Block Design) of the posttest and they did significantly better at the .01 level on one subtest (Comprehension). The disturbed group did significantly better at the .05 level on two subtests (Picture Completion, Block Design) of the posttest and they did significantly better at the .01 level on one subtest (Information). The t-test was then calculated for the three scales (Verbal, Performance, Full Scale). It was concluded that the disturbed group did significantly better at the .01 level on two scales (Verbal, Full Scale) of the posttest. They did not do significantly better upon retest for the Performance scales. The normal group did significantly better at the .01 level on all three scales of the posttest.

The second test of the statistical analysis of the data was analysis of covariance. Analysis of covariance was used to determine if there was a significant difference of IQ scores between the two groups on the three scales on the pre- and posttests. The results showed that there was a significant difference between the adjusted means of the Verbal IQ between both groups at the .01 level. There was no significant difference between the posttest

scores of the Performance IQ between both groups at the .05 level.

CONCLUSIONS

This study concerned itself with determining whether a significant difference would be obtained on test-retest scores obtained with the WISC with institutionalized emotionally disturbed adolescents and with normal adolescents. The analysis of the data concluded that there was a significant difference between the disturbed group and the normal group on test-retest scores of the Verbal scale and the Full Scale, but there was no significant difference found on the Performance Scale.

From the analysis of the subtests of the disturbed group, it can be seen that a significant difference was found on only one Verbal subtest, Information, but there was a significant difference found for the two subtests of the Performance IQ. When looking at the analysis of the scales of the disturbed group, there was a significant difference found for the Verbal IQ and for the Full Scale IQ, and no significant difference found for the Performance IQ. It seems somewhat contradictory that there was a significant difference found on both Performance subtests, when taken separately, and no significant difference found for the Performance IQ, when taken together. The reason for this discrepancy may be attributed to the small sample size. Much of the literature suggests that a significant difference on retest scores could exist on the Performance IQ, but this was not so in this study. One of the reasons for this may be because many of these

adolescents were on anxiety-relieving medication. The Performance section of the WISC is generally influenced by anxiety, since all of the tests are timed and require concentrated effort. Because the medication could not have been discontinued in order for the examiners to test the patients, it is believed that it might have affected the scores. Practice effects should have influenced the scores, but they obviously did not on the Performance section.

Four of the correlation coefficients for the subtests of the disturbed group may be considered as a moderately high degree of relationship. One subtest (Comprehension) showed just a moderate degree of relationship. It is believed that all of the subtests can be accepted with full confidence. The correlation coefficients of the three scales are considered a moderately high degree of relationship and can be accepted with full confidence for use with disturbed adolescents.

From the analysis of the subtests of the normal group, it can be seen that there was a significant difference found on all five subtests between pre- and posttests scale scores. It was concluded that the normal group did significantly better upon retest on all three scales. The literature suggests that the scores of normal children remain somewhat constant on retest, but this study does not seem to agree. Some posible explanations for this lack of consistency in retest scores for this normal group are that they might not have been a good representation of the population, or the sample size was too small. Another explanation might be that adolescence, in general, is a very unstable period of development. During this period, children are maturing into

adulthood and consequently are experiencing anxiety and frustration over this change. The failure of these adolescents to score consistently on a standardized test of intelligence may indicate a changing personality and value system.

All of the correlation coefficients for the five subtests and for the three scales are considered a moderately high degree of relationship. It is believed that the five subtests and the three scales can be accepted with full confidence when used with normal adolescents.

Through analysis of covariance, it was determined that there was a significant difference between the pre- and posttest IQ scores of the two groups on the Verbal IQ and the Full Scale IQ. Much of the literature suggests that the Verbal IQ and the Full Scale IQ are very stable, but the results of this study are in direct discord with this. There was no significant difference found between them on the Performance IQ. The mean IQ of the disturbed group is also much lower than the mean IQ of the normal group. A possible explanation of why the disturbed group had a lower mean IQ than the normal group was because of the emotional problems that these adolescents were having. When an individual becomes mentally ill, the confusion brought about by the illness will usually hinder that individual from scoring as well on an intelligence test as he did before he became ill.

The following is a summary of the conclusions:

1. For the disturbed group, there was a significant difference on test-retest scores on Information, Picture Completion,

and Block Design subtests, and no significant difference on the Comprehension and Arithmetic subtests.

2. For the disturbed group, there was a significant difference on test-retest scores on the Verbal Scale and the Full Scale, but no significant difference on the Performance Scale.

3. For the normal group, there was a significant difference on test-retest scores on all five subtests.

4. For the normal group, there was a significant difference on all three scales.

5. There was a significant difference between the disturbed and normal groups on the Verbal and Full Scale IQ's, but not on the Performance IQ's.

This study concluded that the use of the WISC with emotionally disturbed adolescents is just as reliable as with normal adolescents on the Performance Scale but it is questionable on the Verbal and Full Scales.

RECOMMENDATIONS

The use of psychological tests in diagnosis and remediation is still in its early stages. These tests are very delicate instruments which, when used properly, can be very effective. In order for them to be used properly it is necessary for the examiner to have a thorough background in psychological testing and total familiarity with the test being used.

The WISC is a relatively new test and many of the shortcomings in it are still vague. This study was an attempt to uncover one of these shortcomings and to help psychologists understand

the test better when used with emotionally disturbed children. When the WISC is administered to any adolescent, whether he be exceptional or normal, the psychologist who interprets the results should be aware that these results are only an estimate of the adolescent's intellectual abilities at that time.

So that further research with the WISC for use with emotionally disturbed children can avoid some of the limitations of the present study, the following recommendations are suggested:

1. The sample size should be a more adequate size, since there was an inconsistency in the statistics due to the size of the sample. Also this would be more effective because a larger sample would be more representative of the population.

2. It is suggested that the full test be given to each subject rather than a short form. Vocabulary is known to be the best subtest on the WISC, and the short form which was used in this study did not include it.

3. All subjects should receive the same amount of time between test and retest. It was impossible, in this study, to retest each individual within the same amount of time as all the other subjects because many of the adolescents at the hospital went home on visits. Ideally, the interval time should be between one and two weeks. Most of the subjects in this study were retested within that time.

4. It is believed that the amount of time that the disturbed adolescent had been ill might have affected the results. For future studies on this subject, the researcher should attempt to gather this information.

BIBLIOGRAPHY

BIBLIOGRAPHY

American Psychiatric Association. Diagnostic and Statistical Manual of Mental Disorders, second edition. Washington, D.C.: American Psychiatric Association, 1968.

Anderson, Barry F. The Psychology Experiment. California: Brooks/ Cole Publishing Company, 1971.

Bishop, Dorothy C. More Than One Hundred Years of Concern for the Mentally Ill. Osawatomie State Hospital: Public Information Office.

Buros, Oscar. The Seventh Mental Measurements Yearbook, Volume 1. New Jersey: Grayphon Press, 1972.

Coleman, James C. "Stability of Intelligence Scores in Learning Disorders." Journal of Clinical Psychology, 19, July, 1963, 295-298.

Freeman, Frank. Theory and Practice of Psychological Testing, third edition. New York: Holt, Reinhart and Winston, 1962.

Freidman, Ronald. "The Reliability of the WISC in a Group of M. R. Boys," Journal of Clinical Psychology, 26(2), April, 1970, 181-182.

Garrett, Henry E., and R. S. Woodworth. Statistics in Psychology and Education. New York: David McKay Company, Incorporated, 1966.

Glasser, Alan J., and Irla Lee Zimmerman. Clinical Interpretation of the Wechsler Intelligence Scale for Children. New York and London: Grune and Stratton, 1967.

Irwin, Dale O. "Reliability of the WISC." Journal of Educational Measurement, 3, Winter, 1966, 287-292.

Littell, William. "The WISC: Review of a Decade of Research." Psychology Bulletin.

Palkes, Helen, and Mark Stewart. "Intellectual Ability and Performance of Hyperactive Children." American Journal of Orthopsychology. 42, January, 1972, 35-39.

Quereshi, Mohammed Y. "Practice Effects on the WISC Subtests and IQ Estimates." Journal of Clinical Psychology. 24, January, 1968, 79-85.

Robinson, Herbert B., and Nancy M. Robinson. The Mentally Retarded Child. New York: McGraw-Hill Book Company, 1965.

Roscoe, John T. Fundamental Research Statistics for Behavioral Sciences. New York: Holt, Reinhart and Winston, Incorporated, 1969.

Rosen, Marvin, Linda Stallings, Lucretta Floor, and Myra Nowakiwska. "Reliability and Stability of Wechsler IQ Scores for Institutional Mental Subnormals." American Journal of Mental Deficiency. 73, September, 1968, 218-225.

Solomon, Gerald. "Guidelines on the Use and Medical Effects of Psychostimulant Drugs in Therapy." Journal of Learning Disabilities. 35, November, 1971, 6-11.

Telford, Charles W., and James W. Saurey. The Exceptional Individual. New Jersey: Prentice-Hall, Incorporated, 1972.

Turner, R. K., A. Mathews, and S. Rachman. "The Stability of the WISC in a Psychiatric Group." British Journal of Educational Psychology. 37, June, 1967, 194-200.

Walker, Kenneth P., and Fredrick L. Gross. "IQ Stability Among E. M. R. Children." Training School Bulletin. 66(4), February, 1970, 181-187.

Weschler, David. Manual for the WISC. New York: The Psychological Corporation, 1949.

_____. The Measurement and Appraisal of Adult Intelligence. Baltimore: The Williams and Wilkins Company, 1958.

Whatley, Ruth G., and Walter T. Plant. "The Stability of WISC IQ's for Selected Children." Journal of Psychology. 44, 1957, 165-167.