Fall 8-1966

A CONCURRENT VALIDITY STUDY BETWEEN THE DIFFERENTIAL APTITUDE TEST FORM L AND THE WECHSLER INTELLIGENCE SCALE FOR CHILDREN

Loren Paul Anderson, Jr.
Kansas State College Of Pittsburg

A CONCURRENT VALIDITY STUDY BETWEEN THE DIFFERENTIAL APTITUDE TEST FORM L AND THE WECHSLER INTELLIGENCE SCALE FOR CHILDREN

A Research Project Submitted to the Graduate Division in Partial Fulfillment of the Requirements for the Degree of Specialist in Education

By

Loren Paul Anderson, Jr.

KANSAS STATE COLLEGE OF PITTSBURG

Pittsburg, Kansas

August, 1966

ACKNOWLEDGMENTS

In an endeavor such as this, there were people who helped to make this possible and the writer was indebted to them.

First, to the Advisory Committee, and expecially to Dr. R. E. Galloway, for their time, patience and guiding hand in completing the Project.

Secondly, to the administrative personnel of Washington District Schools and to the staff and students of Coronado Junior High School for their encouragement and assistance in making the research to be accomplished.

Thirdly, to Sister Victorine for editing the paper and her helpful suggestions, and to Don Dibble for the many hours he spent in typing the thesis.

Lastly, to my wife Nita and four children: Linda, Gary, Keith and Cynthia who have been willing to sacrifice in order that the writer might complete the degree.

TABLE OF CONTENTS

LIST OF FIGURES

CHAPTER I

INTRODUCTION TO PROJECT

The purpose of this project was to study the relationship that existed between the scores of fifty boys and fifty girls of a Midwestern suburban junior high school on the Differential Aptitude Test Battery Form L--DAT, a test for measuring the abilities of boys and girls in grades eight through twelve--and the scores of the same students on the Wechsler Intelligence Scale for Children--WISC, an individual intelligence test for children and young people between the ages of five and fifteen. This study had for its ultimate goal a better understanding of the concurrent validity between these two testing instruments. Concurrent validity "is concerned with the relation of test scores to an accepted contemporary criterion of performance on the variable which the test is intended to measure" (Educational Testing Service, 1961).

I. RATIONAL OF PROJECT

Many times educators have requested the results of an individual intelligence test when a group test of scholastic aptitude might have been just as valuable for most practical purposes. In a school of the author's acquaintance, the staff frequently asked the trained psychometrist to administer an individual intelligence test, such as the Wechsler

Intelligence Scale for Children, to help determine the tested intelligence of some of the students. This was a rather time-consuming effort when it was done too often. In many public secondary schools the function of administration of individual intelligence tests falls upon the shoulders of the guidance counselor. The time required in such a case might work a real hardship. Such a problem led to the investigation reported herein.

The Bible related in John 8:32 "and you will know the truth and the truth will make you free" (Bible, 1952). If man was searching for truth in order not to be enslaved, then he must encounter various situations that will bring about conclusions either true or false. Truth is partially found on the printed page. Because of the vast amount of testing that has been done, books, pamphlets, and periodical articles have been written concerning this subject and truth may be drawn from these sources; nevertheless, fallacious assumptions have been drawn by some individuals from a vast amount of data derived through testing.

Schools, industries, and businesses administer tests for a multiplicity of reasons. Much time and expense is involved which needs to be justified not only by the usefulness of the data but also by the reliability, validity, and acceptability of the instruments used. When a test which satisfies these needs has been administered and the data has been recorded, it would appear wasteful of time and money to administer a similar one unless for supportive

reasons. Additional cost in time and money needs to be justified expecially in schools which are financed by taxation. In this study, testing for school purposes was the only focus of attention.

Many counselors had large counselor-counselee ratios, and thus, their responsibilities were many and varied. A counselor with training in psychometric techniques should be able to educate other members of the professional staff in test administration of group tests, thereby, relieving the counselor to function in his unique capacity, that of counseling. It would appear necessary that a school district utilize the leadership and training of the professional counselor in such a way that he could be of service to the most people.

Administering an aptitude test battery to a group would be less time-consuming than administering individual intelligence tests to each member of that group. If therefore, the former would produce results similar to the latter, it would be to the best interest of the school and especially of the counselor to determine the degree of relationship of the two instruments. If the relationship appeared to be high, then it would be apparent that the aptitude test scores would suffice and an individual intelligence test would be unnecessary.

Other considerations would include the greater amount of time and expense involved in the administration of an individual intelligence test, as well as the different

aspects of abilities which would be uncovered by a scholastic aptitude test battery.

To put it more precisely, it would appear unnecessary to give an individual intelligence test to several hundred students of a particular ninth grade class--which would require approximately one hour apiece or hundreds of hours to administer--if a group aptitude test could be administered in five hours to the same number of students, and similar results be obtained.

Another question was raised at this point. Sometimes educators, parents, and even young people themselves belittle the intelligence of junior high school boys and girls in uncomplimentary, if not abusive, language. Since this happens, even occasionally, a research project which would determine whether or not these allegations were true, would seem to be of service.

The purpose of this study was two-fold: (1) to indicate the relationship of a particular individual intelligence test (WISC) with a particular group scholastic aptitude test (DAT) in order to determine if there was similarity in test results, and incidentally, (2) to indicate whether junior high school students in the ninth grade at Coronado Junior High School were less intelligent on the average than other young people of their own grade in the United States.

II. LIMITATIONS OF PROJECT

In any undertaking such as this there were certain limitations. This project was no exception, but mainly they were human and economic--economic in regard to time and money.

In the test administration of the group aptitude test, girls were crowded at lunch tables in a multi-purpose room which was not the most conducive climate for testing purposes. The boys were tested in individual classrooms under the supervision of classroom teachers, some of whom had little instruction in testing procedures.

It would have been impractical to test all 376 ninth graders at Coronado Junior High School with an individual intelligence test; this would have entailed approximately 376 hours of testing time which few school districts would deem feasible.

The time of the counselor who would have been expected to administer the individual intelligence test would have been prevented from his main function as counselor--individual and group counseling. If individual and group counseling were to continue as expected, the services of additional professional staff would be necessary. Though additional testing materials would be involved, the time element would be of greater consequence.

There was another kind of limitation in this project which seemed worthy of attention. Although Coronado Junior

High School, where the study was made, appeared to be a fairly typical Midwestern suburban junior high school, it had its own unique character.

Adler, Murray and Allport have related the uniqueness of the individual because of varying heredity and environmental influences such as experiences and background (Hall and Lindzey, 1957). Likewise, every junior high could be considered unique because its students, faculty, and administrative personnel all have unique, distinctive backgrounds.

The entire individual testing for this project was accomplished by the same individual. It is possible that the examiner could have become lax in the administration and scoring of the test due to the number of tests given over a short period of time. If another person had done some of the testing, this possible source of error could have been reduced and increased the credibility of the results.

Despite the apparent limitations there were indications that the study had sufficient strength to warrant such an undertaking.

III. DEFINITION OF TERMS

In order to understand the terminology and concepts used in the paper, it would be well to define some of the terms. The author will tend to rely on the definitions of the terms used by the authors and publishing houses of the two instruments that were used in the investigation.

Intelligence

Writers down through history have tried to isolate and define the concept of intelligence.

Thorndike et al. related that the obvious hypothesis often advanced was that "intellect is the ability to learn" (Thorndike, et al., 1927). Thorndike was the first to develop clearly the idea that the measurement of intelligence consisted essentially of an evaluation of mental productions in terms of a number and the excellence and speed with which they were effected. Abilities were mental products arranged in different classes of operation. Formerly, psychologists were inclined to use a relatively small number of such classes based mainly on the kind of mental process supposedly involved; psychologists now have altered their classifications to include subdivisions based on material content or factorial analyses. For instance, they subdivided reasoning into three categories: abstract, verbal or arithmetical (Wechsler, 1958).

The Journal of Educational Psychology in 1921 published a symposium on the nature of intelligence. Some of the definitions given by various psychological and educational leaders of that era were:

1. An individual was intelligent in proportion as he was able to carry on abstract thinking - Terman

2. An individual possessed intelligence in so far as he had learned or could learn, to adjust to his environment - Colvin

3. Intelligence seemed to be a biological mechanism by which the effects of a complexity of stimuli were brought together and given a somewhat unified effect of behavior - Peterson

4. Intelligence was an acquiring capacity - Woodrow

5. Intelligence was a general capacity of an individual consciously to adjust his thinking to new requirements - Stern

6. Intelligence was completeness of understanding, inventiveness, persistence in a given course, and critical judgment - Binet (Boynton, 1933).

From this list it could be ascertained that there was anything but unamity of agreement on the definition of intelligence over forty years ago.

More recently some psychologists have been subdividing intelligence rather than making a concrete statement about it. Thorndike suggested that intelligence be subdivided into three main types: (1) abstract or verbal intelligence, involving facility in the use of symbols; (2) practical intelligence, involving facility in manipulating objects; and (3) social intelligence, involving facility in dealing with human beings. This classification emphasized what a person could do, as well as how he could do it (Wechsler, 1958).

The rating which an individual attained on an intelligence test depended on the kind of test used. His score on a test made up largely of verbal items might differ significantly from that obtained on a test of social comprehension

or psychomotor reactions and the perception of spatial relationships (Wechsler, 1958).

Wechsler (1958) related that the opposite could be also true; larger numbers of individuals examined with a variety of intelligence tests who would tend to score high, average or low on one test would do likewise on the rest. This dualism of human abilities, specificity and interdependence, had been reconciled through factor analyses.

The English psychologist, Charles Spearman, made a most profound contribution in this area. He introduced a method for accounting for the variance between paired sets of correlated measures, and also attempted to show by this method that all intellectual abilities could be expressed as functions of two factors, one a general or intellectual factor (g) common to each ability, and a specific factor (s) which was specific to any particular ability and in all cases different from that of all others. Spearman's original methods of factoring a correlational table and his unifactor theory have been largely abandoned by psychologists today; even so, the existence of at least one pervasive factor in all performances requiring intellectual ability has remained a great discovery of psychology today (Wechsler, 1958).

Alexander in his monograph Intelligence, Concrete and Abstract went about testing experimentally Spearman's two factor theory and the unique traits theory, according to which intelligence involved several abilities, each

independent of one another. His experimental study attempted to determine whether test results supported the views that practical and verbal intelligence were each distinct and independent capacities, or the view of Spearman that both were essentially the same in that they were not independent capacities but differed with respect to their specific factors.

Alexander's findings confirmed Spearman's main contention that there was only one common factor in all measures of intelligence, and this factor alone was not sufficient to explain the total correlational variance which existed between the tests used to measure intelligence.

Alexander's investigation also demonstrated that there were factors, such as the subject's interest in doing the tasks set, his persistence in attacking them and his desire to succeed, which might be described as personality factors, but which were recognized in all actual measures of intelligence; they might be appropriately referred to as the non-intellective factors in general intelligence (Wechsler, 1958).

Wechsler indicated that the quantity which we were able to measure by intelligence tests was not a simple quantity. He contended that it was not something which could be expressed by one single factor alone; it was the ability to utilize this energy in contextual situations.

> To concede as much is to admit that any practical definition of intelligence must be fundamentally a biological one in the widest sense

> of the term. That has been the hypothesis assumed in the construction of the author's intelligence scale (Wechsler, 1958).

He further acknowledged that the only thing one could ask of an intelligence scale was that it measure sufficient portions of intelligence to enable us to use it as a reliable index of the individual's global capacity (Wechsler, 1958).

Wechsler (1958) felt several postulates were important: (1) intelligence was a complex function; (2) intelligence was of the nature of a resultant effect; and (3) this resultant effect depended upon the interaction of a theoretically infinite but practically limited number of qualitatively different but additive components. The factors made themselves known in different forms of behavior. A segment of behavior that was factorially defined constituted an ability; these segments of behavior might be grouped into broad classifications as verbal, spatial, numerical, etc.

An intelligence test, according to Wechsler, was one which would seek to evaluate a fragment of behavior in so far as it might be called intelligent. He stated an effective test of intelligence should be made up of tasks calling upon many abilities because people express themselves in different ways.

The thing that was being sought in measuring intelligence was the net result of the complex interaction between the various factors entering into intelligent behavior. An intelligence scale was an assembled battery of tests of

ability; the intelligence rating obtained from them was a numerical expression of their combined contribution.

Aptitude

Drever's Dictionary of Psychology simply defined aptitude as "Natural ability to acquire relatively general or special types of knowledge or skill" (Drever, 1964). Harriman (1961) explained an aptitude as the likelihood for future success usually after training, in a given field of endeaver.

The authors of the Differential Aptitude Tests tended to accept the definition of aptitude from Warren's Dictionary of Psychology. This briefly stated an aptitude as a set of characteristics regarded as symptomatic of an individual's ability to acquire with training some (usually specified) knowledge or skill such as the ability to speak a language (Bennett et al., 1959).

This definition avoided the fallacy that aptitudes were hereditary. Aptitudes were the result of interaction of heredity and environment. Individuals were born with certain potentialities and began to learn immediately.

Aptitudes as defined by Warren's Dictionary implied a very broad base; it embraced any characteristic which was predisposed to learning including intelligence, achievement, personality, interest, and special skills (Bennett et al., 1959).

Bennett, Seashore and Wesman indicated that aptitude could succinctly be defined as "a capacity to learn" (Bennett et al., 1959).

If an aptitude could be defined as capacity to learn, what would an aptitude test propose to do? The 1966 edition of The World Book Encyclopedia defined aptitude as a person's ability or capacity to learn, or capacity to learn with training in certain skills or knowledge.

An aptitude test, as such, could not predict success or failure in a subject; it showed whether a person had the ability to succeed if he had the training and the desire (World Book, 1966).

Thorndike and Hagan indicated that an "aptitude test undertakes to measure what a person could learn to do" (Thorndike and Hagan, 1961).

During the last forty years there was a growing recognition of the need for measuring the many aspects of mental ability. The studies and investigations of Thorndike, Kelley, Spearman, Thomson, Thurstone and others (Bennett et al., 1959) made us aware that intelligence was composed of many abilities which were present in different people in various amounts. A student could have strengths in certain abilities, but limitations in others. The reporting of one total score would obscure the potentiality of the individual.

Industry early recognized the need for aptitude tests; they included such tests as the Stenquit's Mechanical Aptitude Tests, Thurstone's Examination in Clerical Work, the Minnesota Paper Form Board, and the Minnesota Vocational Test for Clerical Workers. Many of the more prominent aptitude tests in recent years sampled narrow segments of

aptitude with the intent of predicting success in a relatively narrow field of vocational endeavor. Bennett, Seashore and Wesman (1959) indicated their aptitude tests were developed to assist the counselor in a look at career possibilities.

Due to the many aspects of ability which aptitude tests now attempt to measure, the task of the school guidance counselor must be enlarged in scope. He must be able to survey the skills and abilities of his counselees in order to arrive at judgments as to which are likely to be most promising.

IV. POINTS OF STUDY

The following points were investigated in this study: (1) whether there was a high correlation between the Wechsler Intelligence Scale for Children Full Scale and the Differential Aptitude Tests--Verbal Reasoning plus Numerical Ability score--for the ninth grade students; (2) by adding the Abstract Reasoning scores with the Verbal Reasoning and Numerical Ability scores, whether there was a higher correlation with scores on the Wechsler Intelligence Scale for Children Full Scale for the hundred boys and girls participating in the study; and (3) whether or not, as a result of an individual intelligence test, the intelligence of one hundred ninth grade students of Coronado Junior High School who participated in the study was average according to the publisher's norms.

CHAPTER II

REVIEW OF THE LITERATURE

A review of the literature was conducted to determine what related studies had been undertaken concerning the validity of individual intelligence tests and scholastic aptitude tests. It was important to determine if there had been any previous studies concerning the Differential Aptitude Tests and Wechsler Intelligence Scale for Children, and if so, to investigate the methods and findings of these studies.

Since the inception of the Differential Aptitude Tests (DAT) and Wechsler Intelligence Scale for Children (WISC), there have been published and unpublished studies, relating the contribution of these instruments to the measurement of scholastic aptitude.

The Educational Index and The Readers' Guide to Periodical Literature were the prime sources used to locate materials and studies for this paper. The DAT Manuals (first and second editions) were borrowed from the Testing Bureau of Kansas State College, and the third and fourth editions were furnished by Psychological Corporation.

Probably the most important question to ask about a test which is being considered for use is: Is the test valid? Jordan stated, "A test is valid in proportion as it measures well what is desired to be measured." (Jordan,

1953). External evidence of validity is found by comparing a test with other measures of the same traits. (Jordan, 1953). A review of such studies follows.

The only published study of concurrent validity that could be located by this writer, correlating the DAT and Wechsler scales, was a study of Federal Reformatory inmates reported in the Journal of Counseling Psychology. Prisoners from three institutions with average mean tested intelligence were given the DAT Verbal Reasoning and Numerical Ability Tests, and the Wechsler Adult Intelligence Scale in 1957. Ninty-eight 16 and 17 year olds, sixty-one 18 and 19 year olds and ninety-five 20 and 24 year olds were tested. Correlations between the DAT VR+NA and the WAIS Full Scale I. Q. were as follows: for the 16 and 17 year olds, .79; for the 18 and 19 year olds, .75 and for the inmates 20 to 24 year olds, .74 (Doppelt et al., 1959).

Supplementary validation and related studies concerning the Differential Aptitude Tests and the Wechsler Intelligence Scale for Children are presented below.

I. VALIDATION STUDIES OF DAT

There was no mention in the Manual of the use of factor analysis to assure validity in the development of the DAT, so recourse was had to other sources to determine whether or not it was used. An article by Bennett, Seashore and Wesman (1948) gave the following reasons why the factorial analysis was not used as a direct basis for the development of the

test battery: (1) there was evidence that the factors isolated on the basis of data from one sample were quite different from those obtained from another sample; (2) experience indicated that several tests were usually required to account for the variance of each factor; and (3) factor scores were not readily interpreted by persons working in the field of guidance (Bennett et al., 1948).

Cottle (1948) indicated that the authors of the DAT felt that the tests should be used for predictive purposes early in grade eight, nine or ten, but their usage in an attempt to show validity of the test in this case would be measuring proficiency and not prediction. In Section E of the Manual, "Interpretation of the Tests," the DAT writers stated: "To the extent that some of these should be associated with school success the coefficients are evidence of validity" (Bennett et al., 1947).

The authors of the DAT in the same Manual related that in the development of the test, pending the publication of statistical and experimental reports, interpretation of test scores would depend upon the judgment of the authors and their associates and upon the expertness of the educators and psychologists who would use the tests in practical counseling situations. They stated further that enough was known about tests to permit such arguments by analogy. The tests of the DAT were recognized as being similar to a variety of psychological tests in purpose and general content. The test publisher admitted that analogous reasoning

with regard to the validity of one test, though it was essentially similar with other tests, could lead to some errors, and was a defensible approach only on a temporary basis until such time when the test authors and publishers would be committed to a specific program for replacing validation by inference with validation by statistical and experimental data (Bennett et al., 1947).

The writers of the DAT contended that even though the DAT was released with validity coefficients almost absent, the battery had its roots from previous research such as that on the Mechanical Reasoning Test which was a direct outgrowth of the Bennett Tests of Mechanical Comprehension (Bennett et al., 1947).

The counselor could only guess at the validity when the DAT was first published. It was necessary that caution be exercised in calling conclusions drawn from their use and that further information would be gained only through research and the conscientious reporting of this research in the literature (Cottle, 1948).

Bennett, Seashore and Wesman (1947) in Section E of the DAT Manual titled "A General Guide to the Interpretation of the Test" stated: "Thus, the Verbal Reasoning, Numerical Ability and Abstract Reasoning Tests measure those functions which are associated with 'general intelligence.'" They added that the Verbal and Numerical Tests alone would tap the same general area as tests which had no nonverbal content, and the Abstract Reasoning Test tapped another aspect

of "general intelligence," providing a measure of reasoning ability (Bennett et al., 1947). Yet as of 1948 there was no proof to substantiate this.

The first correlational studies of the DAT tests and intelligence and general aptitude tests appeared in 1951. Correlations were obtained between the DAT and Henmon-Nelson I. Q. The correlations between the DAT tests and Henmon-Nelson for sixty-eight boys were as follows: Verbal Reasoning .73, Numerical Ability .70, and Abstract Reasoning .52; for fifty-eight girls correlations of .77, .75 and .71 respectively on the same tests (Bennett et al., 1947). For the Otis Gamma Q-S for 106 boys: correlations of .73 for Verbal Reasoning, .51 for Numerical Ability, and .57 for Abstract Reasoning were obtained; for 136 girls .75, .67 and .47 respectively on the same tests (Bennett et al., 1947). As a result of this and other data the publisher of the DAT stated:

> We may safely generalize that the DAT will correlate well with almost any of the standard 'intelligence' or scholastic aptitude tests--in some instances about as well as the 'intelligence' test would correlate with an alternate form of itself (Bennett et al., 1947).

The authors of the DAT indicated the use of intelligence tests was unnecessary where the DAT was already being used (Bennett et al., 1947).

The DAT Manual (third edition) stated:

> A large body of experimental evidence substantiates the belief that the DAT Verbal Reasoning and Numerical Ability do, in fact, measure what is measured by intelligence and

> scholastic aptitude tests and are effective predictors of future academic performance (Bennett et al., 1959).

Some of the correlations that have been reported since 1952 in further establishing this was presented in table form in this Manual (Bennett et al., 1959).

Bennett, Seashore and Wesman (1963) appeared to have felt confident about the validation studies of Form A and B; for they implied in a current pamphlet that Verbal Reasoning and Numerical Ability were a suitable combination for secondary purposes, and abilities measured by these tests have been represented in intelligence tests more commonly called scholastic aptitude tests.

The Fourth Manual for the Differential Aptitude Tests (Forms L and M) presented some coefficients of correlation between the DAT Forms L and M and intelligence and ability tests. The results of this data were from four communities as listed in the Manual.

The DAT Forms L and M and the California Test of Mental Maturity (1957 Short Form Junior High School Level) Total I. Q. produced a coefficient of correlation of .75 for males and .84 for females (Bennett et al., 1966).

A correlation of .78 for males and .79 for females was obtained between the DAT Forms L and M and the Henmon-Nelson Test of Mental Ability (Revised Edition, Form A) (Bennett et al., 1966).

Correlations of .72 to .83 were obtained for males between the DAT Forms L and M and the Lorge-Thorndike Verbal

Intelligence Tests (Form A, Level 4); the correlations for girls were from .82 to .86 using the same test instrument (Bennett et al., 1966).

The DAT Forms L and M and Lorge-Thorndike Non-Verbal produced correlations of .63 to .75 for males, and .66 to .77 for females (Bennett et al., 1966).

A coefficient of correlation of .85 for males and .81 for females was obtained between the Otis Quick-Scoring Mental Ability Tests (Form Gamma FM) and the DAT VR+NA Forms L and M (Bennett et al., 1966).

The DAT VR+NA and the School and College Ability Tests (Form 2A) Total produced correlations of .90 for males and .89 for females (Bennett et al., 1966).

No concurrent validity studies between the DAT Forms L and M and an individual intelligence test were reported in the DAT Manual (fourth edition).

Segel (1947) stated his belief that one of the important steps in the growth of measurement work at the secondary school level would be the use of multiple aptitude tests, and these tests would assist in diagnosing more fundamental strengths and weaknesses than other types of tests. The value of general intelligence tests at the secondary level would become limited as multiple aptitude tests were better developed. The multi-aptitude test would break down the general mental ability score so that the type of traits the student had would be made known.

If one wished to measure both academic ability and special aptitudes, the DAT might well be a wise choice, asserted Layton and Swanson; a study they carried on sought to extend the existing information bearing on the usefulness of the DAT for predicting long-term academic success.

In 1951-52 when the DAT was a part of the Minnesota Statewide Testing Program, about 108 schools administered it to 4,600 students. In 1954 after the same high schools had completed testing their juniors on different tests, 27 of the 108 schools were selected at random. For the 27 schools all students were selected who had a complete set of both DAT scores and junior measures (American Council on Educational Psychology Examination, the Cooperative English Test scores and high school percentile rank). Relationships of high school freshmen test scores to these measures provided an important chain of information about a student for the entire high school career.

There were only two parts of the DAT, Verbal Reasoning and Numerical Ability, that combined significantly in a coefficient of multiple correlation in predicting the academic measures; for boys these coefficients were .69, .67 and .63 for the A. C. E., Cooperative English Test and high school percentile rank respectively. For girls the coefficients were .71, .68 and .61 for the above three scales respectively. Layton and Swanson (1958) contended that anyone who used the DAT should not be surprised that some

parts of the DAT have low relationships with academic measures. The relationships do not invalidate the use of the DAT with those students who do not have high academic promise. The DAT was one set of tests which would indicate to what degree the student possessed special aptitudes upon which he might capitalize in making a vocational choice other than one requiring college.

In recent years aptitude tests have progressed from dependence upon a single test of general intelligence to a battery of tests covering a wide range of specific aptitudes (Flanagan, 1955). Stanley in his review implied that traditional group intelligence measures were being threatened by differential ability batteries, particularly the DAT (Warrington and Saupe, 1959). Correlations between the eight DAT subtests and seven group tests of intelligence were in general so substantial that Bennett, Seashore and Wesman deemed it unnecessary to employ an intelligence test when DAT results were available (Stanley, 1953).

Williams (1952) obtained an r of .73 between the DAT Verbal Reasoning and I. Q.s on Form L of the Revised Stanford-Binet Intelligence Scale for fifty white sophomores; a correlation of .78 for DAT and Henmon-Nelson was also calculated.

Even though some of the newer differential batteries deserved popularity, Warrington and Saupe (1959) testified that the development and use of venerable tests of intelligence continued at the usual rate. They based their conclusion on the case with which references of Stanley's review

were collected. It appeared that publishers of differential batteries, not having been able to supply single test scores, had decided to accept their value. Psychological Corporation in 1958 provided evidence of predictive validity and norms for the sum of DAT scores, Verbal Reasoning and Numerical Ability, used as a single measure of scholastic aptitude.

Psychologists have now identified and defended twenty to twenty-five types of aptitude tests, each of which had sufficient uniqueness in the functions measured to warrant identifying and reporting as a separate score. Flanagan (1960) felt that aptitude tests alone were insufficient evidence to determine appropriate types of careers; inclination and educational background also needed to be considered.

Project Talent was a program to develop a national inventory of the aptitudes of a scientifically selected sample of half million high school students. The students were administered a series of aptitude, achievement, interest and biographical information tests in March, 1960; other data were also collected. Information was made available regarding the success of students having specific aptitude patterns in completing the first year's work in a wide variety of college courses and occupations.

One of the objectives of Project Talent was the development of a set of standards for educational and psychological measurements which would provide more accurate standards for test authors in standardizing tests so that scores would indicate comparable levels of ability (Flanagan, 1960).

Dyer (1960) in his provocative article titled "A Psychometrician Views Human Ability" contended that most people had the superstition that an aptitude test measured native ability. An aptitude test measured the quality of a pupil's performance on a number of mental tasks; it would relate how well an individual could cope with tasks like those on the test at the time he took the test, and he stated, it told nothing more; anything beyond this was interference. Aptitude test scores were inevitably determined by what the child had seen, heard and done.

The psychometrician thought of ability not as a constant entity inside the individual, but merely the quality of his behavior with respect to any set of tasks which may confront him at the moment. Ability was always a construct, derived from the test score; one didn't measure the construct, but one arrived at it through the measurement of behavior. Using this kind of definition, an individual's ability might vary with time and with the kinds of tasks that made up the test.

A shortcoming of multi-aptitude test batteries, according to Dyer (1960), was that they encouraged the oversimplified notion that a profile of ability scores summarized clearly and efficiently a large amount of reliable information about most of the important abilities of an individual. A fool-proof type of profile chart was developed called differential prediction studies; the goal of these studies was to provide a profile of probable success in several

educational or vocational fields rather than a profile showing where a student stood on a series of tests. Dyer did not believe the profile chart provided much differential prediction.

Hills (1964) stated in his overview of scholastic aptitude tests that the points he looked for when reviewing tests included: Was the test the right sort for the purpose intended, such as age level, educational level, and what special kind of person? Was the test essentially a speed or a power test? Did the manual contain sufficient appropriate data and interpretative material of high quality to make the test immediately useful or was it only likely to be an instrument upon which interesting research might be conducted? Hills was primarily interested in a scholastic aptitude test to see if it was valid; if the scores were not valid, high reliability was no help.

The intercorrelations between the individual tests of the DAT taken by 497 eighth grade students, as calculated by Lundy and Shertzer (1963), were as follows: .70 and above for Verbal Reasoning and Sentences; .60 and above for Numerical Ability and Verbal Reasoning, Abstract Reasoning and Verbal Reasoning, Mechanical Reasoning and Verbal Reasoning, Spelling and Verbal Reasoning. This appeared to point up the importance of verbal ability and its close relationship to the individual tests of the DAT and suggested the importance of reading as associated with each of the tests.

II. VALIDATION STUDIES OF WISC

Littell (1960) indicated the concurrent validity of the WISC (Wechsler Intelligence Scale for Children) had been restricted to correlations between the WISC and other measures of intelligence or achievement.

In 1951 Frandsen and Higginson (1951) reported a study of fifty-four fourth grade children, and concluded that I. Q. norms from the WISC and Stanford-Binet were comparable at least within the range of one or two standard deviations above and below the mean; this was the most favorable and unqualified statement of the comparability of the WISC appearing in the literature; the correlations were Verbal .71, Performance .63 and Full Scale .76. Studies by Krugman et al. (1951), Weider et al. (1951), Mussen et al. (1952), and Stroud et al. (1957) produced similar correlations.

Holland (1953), Harlow et al. (1957), Arnold and Wagner (1955) compared performance of the WISC and Stanford-Binet and found that Stanford-Binet scores correlated higher with Full Scale and Verbal Scale than with Performance. They concluded that the WISC was an adequate valid measure of intelligence in the age range six to fourteen years.

The number of studies conducted and the high correlations obtained between the WISC and Binet were impressive. It was likewise impressive that DAT correlations of VR+NA and other ability and achievement measures have produced correlations from .70 to .89 (Bennett et al., 1959).

Somewhat puzzling, however, was the lack of published or unpublished studies of the relationship between the DAT and WISC. Would the authors of the DAT and WISC lead us to believe that one study between the DAT VR+NA and WAIS Full Scale, and the studies of DAT VR+NA and group intelligence and ability tests would be sufficient? It seemed difficult to understand why concurrent validity studies of two reputable measures published by the same company, the DAT and WISC, have not been reported in published literature.

Diamond (1947) took the approach of considering subtests in groups designated as linguistic, clerical and spatial. In a study of a hundred men with average intelligence, there was a high degree of relationship between each of the groups of Wechsler-Bellevue subtests and the aptitude test with which it had been paired in the study; the correlations ranged from .59 to .73. The relationships may be taken as evidence of the validity of the group scores as indicators of linguistic, clerical and spatial aptitudes. Diamond suggested that when the Wechsler-Bellevue was used for vocational guidance purposes, consideration should be given to the advisability of including in the report a statement of the client's relative success in the three subgroups in addition to reporting the Verbal and Performance ratings in the standard manner. The three groups included the following Wechsler-Bellevue subtests: linguistic (Information, Comprehension and Similarities); clerical (Digit Span, Arithmetic and Digit Symbol); and spatial (Picture Completion, Object Assembly and Block Design).

In a study by Burks and Bruce (1955), the WISC was administered to thirty-one poor readers and eleven good readers. The poor readers were significantly low on Information, Arithmetic and Coding, whereas they were significantly high on Picture Arrangement, Block Design and Comprehension. The good readers were significantly high on the Similarities subtest. The investigators hypothesized that the poor readers as a group approach learning situations in a more concrete manner as a result of an inability to handle abstractions.

Estes (1953) made a study of the WISC from a socio-economic approach. Two groups of students were studied: forty from an upper economic status and forty from a low economic status. There were an equal number of girls used in the report; 50% of the students were in the second grade and 50% of the students were in the fifth grade. She concluded that differences in socio-economic level effected differences in scores on the WISC at the second grade, but were not significant for the fifth grade.

There was not a significant difference between the two economic groups as to predictive power of the WISC. Using Fisher's Z transformation, it was found that the predictive relationship with achievement was not significant either at the second or fifth grade. When achievement was used as the criterion, the WISC did not predict better for one economic group than the other (Estes, 1953). Chauncy (1959) related that tests administered in junior high school years were

about as predictive of future success as were tests administered towards the end of the senior high school.

What about the question of anxiety in taking individual intelligence tests? Siegman (1956) administered the WAIS and Taylor Personality Scale of Manifest Anxiety to Thirty-five medical and psychiatric patients. Subjects who received high scores on the Taylor Personality Scale obtained significantly lower scores on the timed than on the untimed subtests of the WAIS; this suggested that anxiety had a disruptive effect.

Malnig (1964) used the basic assumption that under psychological stress there would be greater variability in the performance on the WAIS of subjects with low anxiety. He arranged the scores of 210 male college students in descending order according to Taylor Manifest Anxiety Scale. In his results in no case did the correlations for the high anxiety group attain significance, while for the low anxiety group every correlation was significant at the .01 level of confidence.

In a test-retest situation, Gehman and Matyas (1956) found the mean WISC Performance I. Q. significantly greater than the WISC Verbal I. Q. at the ninth grade level as compared with the fifth grade level; the Verbal I. Q. was 96.72 whereas the Performance I. Q. was 103.23 at the ninth grade level. Sixty-seven percent of the ninth grade pupils had higher Performance than Verbal I. Q.s. The difference of 6.51 I. Q. points was almost identical with a finding from

Delattre and Cole (1952) working with children of ages 10-5 to 15-7 who tended to be above average in intelligence. These discrepancies implied a possibility that the Verbal and Performance Scales were not of equal difficulty at the older age levels. There was another speculation concerning the cause of the discrepancies; possibly the Performance Test was more sensistive to coaching when intensive testing was done in one school system than was true of the Verbal Scale (Gehman and Matyas, 1956).

Using a large experimental population, Hunt et al. (1948) made an investigation designed to test the validity of five previously used abbreviated intelligence scales and to devise some new ones for future use; with a group of high school students they obtained a correlation of .96 between the subtests Comprehension, Arithmetic, Similarities, Picture Arrangement and Digit Span of Wechsler-Bellevue. Wallen (1962) reported in the February 1962 issue of Review of Educational Research that there was an interest in abbreviated forms of the Wechsler tests by combining subtests. Clayton and Payne obtained correlations in the .90's between three test combinations and the total score. Thompson and Finley in 1963 selected subtests from the WISC which would best predict Full Scale I. Q.; from a sample of 151 gifted elementary school children, the scores on their short form correlated .84 with the actual I. Q. scores based on the Full Scale. The predicted score varied only four I. Q. points from the actual score in seventy-four percent of the cases and within

eight points in all but two percent of the cases (Millman and Glock, 1965).

Kureth et al. (1952) made a study to determine the validity of the WISC at the five and six year levels with the Revised Stanford-Binet Form L as a criterion. Of the one hundred children in the study, correlations of the WISC I. Q. with Binet ranged from .71 to .84 and was significant beyond the .01 level of confidence; the order of merit from highest to lowest was Full Scale, Verbal Scale and Performance Scale. Considering the available data one could hardly doubt that in general the WISC was measuring the same thing as the Binet with in these age groups. The Performance Scale did not seem to be measuring a distinct aspect of intelligence since the correlations with the Binet were high and quite similar to those for the Verbal Scale. On the basis of their findings Kureth and her associates generally agreed that the WISC subtests were valid.

Another study on the validity of the WISC was made by Pastovic and Guthrie (1951). Their findings indicated a higher mean Performance I. Q. than the mean Verbal I. Q. over a wide range of intelligence.

Altus (1952) administered the WISC to a group of fifty-five young people in a junior high school in Santa Barbara County, California. Scores were already available from the California Test of Mental Maturity and the Progressive Reading Test. Correlations of the WISC Full Scale I. Q. with the California Test of Mental Maturity Total I. Q. yielded an r

of .81 which was typical of the intercorrelation of accepted intelligence measures.

In the past there has been a tendency to regard some tasks as measures of intelligence and others as measures of aptitudes. Most people now have advanced to the point of thinking in terms of multiple scores to appraise multiple abilities. Perhaps this is because test users have dealt with profiles based on several more or less independent test scores, each with some validity for clinical diagnosis and educational or vocational prediction.

The WISC used a discrepancy score which meant the numerical difference between a child's Verbal I. Q. and his Performance I. Q. Seashore (1951) related in his discussion on the differences between the Verbal and Performance on the WISC that one can expect about one-third of a group of random subjects to show a difference between their Verbal and Performance I. Q.s greater than 12 to 13 points of I. Q., while in about two-thirds of the cases the differences would be greater than this about the same number of times. A person could expect similar differences between Verbal and Performance I. Q. among children at all ages who were examined with the WISC.

In the WISC standardization sample of 2,200 subjects--100 boys and 100 girls at each age from five to fifteen--only 4% showed zero differences between Verbal and Performance I. Q.; about 75% of the subjects showed differences of four points or more; 50% of them eight points or more; and 25%

of them fifteen points or more; and five percent of the children showed Verbal and Performance I. Q.s as great as twenty-five points or more.

> Since tests are fallible, we must recognize that some part, and perhaps a considerable part of the I. Q. discrepancies we observe within the individual are due to errors of measurement (Seashore, 1951).

As a result it has been necessary for the purpose of interpreting test scores to develop and exercise the habit of thinking plus or minus so much when noting or discussing them.

Seashore (1951) reports still other findings from the use of that test instrument. Except for the children whose parents were professional and semiprofessional, there was no serious tendency for children in diverse occupational groups to show relative excellence as measured by the WISC.

III. LIMITATIONS OF THE LITERATURE

The most apparent limitation was the lack of published studies of concurrent validity between the DAT and the WISC; in fact--as previously mentioned--they were not to be found.

Since the DAT Form L was published in 1963, there had been some correlational studies with various group intelligence and ability tests, but even these were limited to four communities.

The authors of the DAT reported numerous studies, but only one concerning the Wechsler Scales, and that one was between the DAT and WAIS of a sample of Federal Reformatory

inmates of high school age (Doppelt and Seashore, 1959) which was probably not a true sample of the general population. There were no additional studies reported in the DAT Manual (fourth edition) between the DAT and Wechsler Scales.

The literature indicated that the DAT and WISC were both valid and reliable instruments of themselves, and tested the properties they were supposed too.

Because of the lack of any additional information in the literature concerning the DAT and Wechsler Scales, the project reported in this paper appeared to be in order.

CHAPTER III

METHOD AND PROCEDURE

A concurrent validity study was undertaken, using the Pearsonian r, between the scores from DAT (Form L) VR+NA and WISC Full Scale scores, DAT VR+NA+AR scores and WISC Full Scale scores for one hundred ninth grade boys and girls from a Midwestern suburban junior high school. The fifty males and fifty females were chosen at random from a class of 376 students to take the individual intelligence test; they had taken the scholastic aptitude test with the rest of the ninth grade students as part of the school's basic testing program.

For the sake of the study, hypotheses were formulated.

I. HYPOTHESES

In making hypotheses in a statistical study of a sample group, the concept of null hypothesis must be understood and employed. The null hypothesis is one which hypothesizes that no real differences exist between several groups studied or between values that represent those groups. (Cottle and Downie, 1960). After the statistical study had been conducted, the null hypotheses were to be either accepted or rejected. Acceptance of the null hypotheses in this study, for instance, would mean that the data indicated there was no significant difference between the two coefficients of

correlation, the one calculated from the data and the other representing no correlation. On the other hand, the null hypothesis would be rejected where the study indicated a real difference existed between the observed coefficient and a zero coefficient.

The null hypotheses in this study were as follows: (1) there was no difference between the DAT VR+NA score and WISC Full Scale I. Q. for the ninth grade students in Coronado Junior High School; (2) there was no difference between the DAT VR+NA+AR scores and the WISC Full Scale I. Q. for those same students; and (3) there was no difference in tested intelligence as measured with the WISC Full Scale I. Q. between ninth grade students of Coronado and ninth grade students in WISC published norms.

II. POPULATION

The population in this validity study included all ninth grade students in Coronado Junior High School (Washington District Schools), Kansas City, Kansas, from 1965-70.

Community

The district called Washington District Schools is located in several political divisions. Over fifty percent of the District lies within the Kansas City, Kansas city limits; this was done through annexation procedures that were accomplished in 1965 and 1966. The rest of the District lies in townships surrounding that city. The townships have

existed to carry out the functional aspects of local government, but have nothing to do with the school district.

The population of the school district was approximately 40,000 with more people moving into the area yearly. The increase tended to come from Kansas City, Kansas, as people purchased new homes. Racial integration in Kansas City, Kansas caused the migration of some Caucasian families to the suburban area.

A majority of the people in the school district were homeowners. The homes ranged from small, poorly constructed dwellings to a few in the $50,000 range and upward; most of the new homes were in the $15,000 to $25,000 range. The rental property consisted mainly of duplexes and other small multiple-dwellings. Because many people were homeowners, any exodus from the school district was mainly to the west of the District, or to the south to parts of Johnson County which was considered more of a prestige locality, or completely out of the state. Very few of the patrons moved to rural areas of Kansas.

Occupational patterns ran the gamut in Wyandotte County. A majority of the men worked at local industrial plants in the metropolitan area, or were semi-skilled or skilled craftsmen; however, there were some who were considered unskilled laborers. The white collar workers were employed mainly in clerical jobs or as sales personnel. There were a few individuals who had managerial or executive positions, but this was the exception, not the rule. Professional people such as

doctors, lawyers, engineers, and teachers comprised a small percent of the working force of the locale. Many women worked because of the cost of living that demanded additional family income; their jobs were primarily clerical, sales personnel, operators in factories, or teachers. A few women owned their own businesses such as beauty shops, gift shops, etc.

A majority of the people were protestants, but there were many Roman Catholics living in the school district. The major Protestant faiths were Baptist, Methodist, Lutheran, Disciples of Christ, and United Church of Christ; there were also a portion who belonged to the Nazarene, Assembly of God, and various Pentecostal churches.

Entertainment and recreational activities were quite diverse. A large county park that provided boating, fishing and picnicking was adjacent to the school district. Kansas City, Missouri afforded entertainment by way of its theatres, zoo, professional ball teams, and other attractions where the people could enjoy their leisure time.

There were organizations in Wyandotte County that had established ball teams of various sorts for children and young people; these clubs were quite active in the summer time in the baseball leagues.

There were several historical museums in the vicinity which interested not only tourists but the area people as well. The Truman Library at Independence, Missouri and the Nelson Art Gallery in Kansas City, Missouri were renowned

centers for sponsored or independent tours. Operas, concerts and exhibitions of various kinds were scheduled regularly.

School District

Since there were no large industrial plants within the area of Washington District Schools, the tax base was primarily from local property owners; this created somewhat of a burden upon the people in the district. There were two shopping centers that contributed to the tax base plus several small businesses, such as beauty and barber shops. Federal funds were limited since very few of the adults work in Federal jobs that are under governmental control. As a result of Federal legislation, money became available to purchase supplies and equipment, thereby permitting local and state funds to be used for salaries and building programs. With the establishment of the state foundation program, more revenue came from the state although not in an appreciably large amount to Wyandotte County at that time; however, it did provide some relief for local property owners.

Washington District Schools until recent years was a multiplicity of school districts composed of a rural high school district and various elementary school districts; the elementary districts tended to operate independently of each other, and had a quasi-independent relationship with the local high school.

In 1959 the people in the school district voted bonds to build two junior high schools. This brought about in 1960

a change in organization of the school district: from the 8-4 plan to the 6-3-3 plan.

The people desiring a uniform educational program voted in November, 1961, to unify the school district. The organizational pattern remained the same except that kindergartens were added in schools that did not have them at the time. The policies were now formulated by one board rather than several; these policies were executed by the superintendent, assistant superintendent, and other administrative personnel hired by the board. As a result of unification, each student would be provided with similar experiences regardless of the school he attended.

Included in the district at the time of this project were eleven elementary schools, two junior high schools, and one senior high school. Two of the elementary schools were opened during the 1965-66 school year. One of the elementary schools was Negro with an enrollment of about eighty students. The other elementary schools' enrollments ranged from sixty to over a thousand students.

The two junior high schools, Coronado and Arrowhead, had student populations of 1,124 and 1,020 respectively. The senior high school had an enrollment of over 1,800 students.

There were two Catholic elementary schools in the District, but about one-fourth of these students entered the public schools at grade nine rather than attending the parochial high school in Kansas City, Kansas.

A change in the structure of the school system can be envisioned. The people in the District voted to attach themselves to the Kansas City, Kansas school district which will probably be completed in July, 1967.

This is a resume of the environmental setting of the project.

School

Coronado Junior High School was one of two junior high schools located in the Washington District Schools. It comprised grades seven through nine. The school during 1965-66 had a professional staff which included a principal, an assistant principal, two counselors, a nurse, a librarian, a reading specialist, and forty-two classroom teachers. In 1965-66 there were 1,124 students in the school.

III. SAMPLE

The number of students (one hundred) were taken from the general population of about 2,000 ninth grade students. If possible it would have been best to use the total ninth grade class in the validity study, but it was neither practical nor feasible.

Because the number of boys equalled the number of girls in the class, results of test data from fifty boys and fifty girls were used.

The sampling was done by a random method. All ninth grade boys enrolled during the 1965-66 school year were

assigned a number and then the numbers were placed in a hat, shuffled several times, and fifty numbers were then drawn. The same procedure was followed for the girls. The conclusions of this study were drawn from the sample chosen to represent the total ninth grade population of the school from 1965 to 1970. This sample constituted about twenty-seven percent of the ninth grade class.

IV. INSTRUMENTS USED IN STUDY

The two test instruments used in the study were the Differential Aptitude Tests Form L and the Wechsler Intelligence Scale for Children.

Differential Aptitude Test Form L

According to the authors of the DAT,

> The Differential Aptitude Tests were developed to provide an integrated, scientific and well-standardized procedure for measuring the abilities of boys and girls in grades eight through twelve for purposes of educational and vocational guidance (Bennett et al., 1959).

The DAT Form L published in 1963 was a revision of the DAT Form A that came out in 1947. Four of the eight tests in Form L were unchanged: Numerical Ability, Abstract Reasoning, Language Usage-Spelling and Mechanical Reasoning; the other four tests which include Verbal Reasoning, Space Relations, Language Usage-Grammar and Clerical Speed and Accuracy had the same kinds of questions and posed the same intellectual tasks as in Form A, but had been improved in format to make them more readily adaptable to modern

automatic scoring machines (Psychological Corporation Test Catalog, 1965).

Schools from ninety-five communities in forty-three states, representing all major geographical areas, contributed to the normative sampling for Forms L and M. More than 50,000 students of grades eight through twelve from 195 schools were included in the standardization sample. The testing in the normative sample was done during September, October and November, 1962; this represented the first semester norms. The authors of the DAT implied that scores increased over the course of a year, and so second semester norms were obtained by interpolating between the corresponding entries of successive grades tested in the fall standardization program (Directions for Administration and Scoring and Norms Forms L and M, 1963).

The 1960 United States Census served as the basis for determining the regional breakdown of school enrollees from communities of various sizes within each region. For instance, the publishers wanted to include 28.4% of their sampling from the North Central States; in their grade nine sample they were able to obtain 29.2% of the boys and 29.9% of the girls from this region.

In the DAT Form L and M standardization by community size for urban fringe areas, they desired 5.5% of the total sampling. For grade nine they tested 2.7% of the boys and 3.0% of the girls which was comparable to what they hoped

to test for urban fringe areas (Directions for Administration and Scoring and Norms Forms L and M, 1963).

Because of the sex differences in abilities as measured by the DAT, a separate set of norms for boys and girls were provided.

The tests of the DAT were timed. The complete battery would be usually given in two to six testing sessions. The testing time allowed for the complete battery of the DAT Form L was 179 minutes (Directions for Administration and Scoring and Norms Forms L and M, 1963).

The Verbal Reasoning Test was constructed as a measure of ability to understand concepts framed in words; it tried to evaluate the student's ability to generalize and to think constructively, rather than to demonstrate simple vocabulary recognition.

The Numerical Ability Test was designed to test understanding of numerical relationships and facility in handling numerical concepts. The problems were framed in the item type called arithmetic computation rather than arithmetic reasoning. The desire was to avoid the language elements of the usual arithmetic reasoning problem in which reading ability might play a significant role.

The Abstract Reasoning Test was devised as a non-verbal measure of the student's reasoning ability. This test supplemented the general intelligence aspects of the Verbal and Numerical Tests. It involved the ability to perceive

relationships in abstract figure patterns. This test could not serve as a substitute for the Verbal Reasoning Test, but might be valuable as a check on the Verbal score in some cases of known or suspected language handicap (Bennett et al., 1966).

Wechsler Intelligence Scale for Children

The WISC was an outgrowth of the Wechsler-Bellevue Form II. New items were added at the easier end of each subtest to permit examination of children as young as five years of age.

The WISC was constructed to measure intelligence as defined by Wechsler in the first chapter of this thesis. It was standardized with exceptional care over a five-year period of experimental tryouts, field testing and statistical analysis.

The WISC was standardized on a sample of 100 boys and 100 girls at each age level from five through fifteen years; each child was tested within one and one-half months of his mid-year. A total of 2,200 Caucasian boys and girls were tested. The test could be administered to individuals between the ages of 5-0 and 15-11.

The North Central States contained 32.7% of the general population at that time whereas Wechsler used only 28.9% from this area in standardization. Bedroom villages attached to large cities were generally avoided in the standardization sample (Seashore et al., 1950). This would indicate that

suburban areas, such as involved in this study, were not well represented in the standardization process.

During the standardization process for the WISC, Verbal subtests correlated .67 with the Verbal Score, .46 with the Performance Score and .61 with the Full Scale Score. The Performance subtests correlated .42 with Verbal Score, .51 with Performance Score and .52 with Full Scale Score. The Verbal tests correlated higher with the Verbal Score than with the Performance Score. The Performance tests correlated higher with the Performance Score than with the Verbal Score. The correlations between the Verbal Score and the Performance Score were sufficiently high with .60, .68 and .56 for ages 7½, 10½ and 13½ respectively which indicated a considerable common variance; yet they were low enough to suggest that the abilities included in Verbal and Performance tests could not be readily inferred from each other. Both areas of abilities need to be tapped in an over-all appraisal of abilities (Seashore et al., 1950).

The standard errors of measurement in I. Q. units at 7½, 10½ and 13½ respectively were as follows: Verbal, 5.19, 3.00 and 3.00; Performance, 5.61, 4.98 and 4.74; and Full Scale, 4.25, 3.36 and 3.68 (Seashore et al., 1950). Although these statistics were not deemed necessary for the interpretation of data in this study, the information was included in this report in order that complete details of the WISC instrument would be included.

The Verbal subtests of the WISC included Information, Comprehension, Arithmetic, Similarities and Vocabulary; Digit Span was a supplementary subtest. The Performance subtests were Picture Completion, Block Design, Picture Arrangement, Object Assembly and Coding; the Mazes subtest was a supplementary Performance test. Speed was a factor for Arithmetic, Picture Completion, Block Design, Picture Arrangement, Object Assembly, Coding and the supplementary subtest Mazes; the other subtests were not timed (Wechsler, 1949).

The instrument was administered individually and took about an hour to give.

V. STATISTICAL TECHNIQUES

The WISC was administered to one hundred ninth grade students at Coronado Junior High School by this writer during the 1965-66 school year. The protocols were randomly reviewed by two school personnel who had had training in administering and scoring the WISC; they were Charles Byler, assistant principal at Coronado Junior High School, and Fred Madaus, counselor at Arrowhead Junior High School (Washington District Schools).

The DAT Form L was administered on October 12 and 13, 1965, to all ninth grade students. Boys were given the test battery in classrooms. One hundred eighty-two girls were tested at one time in the school's multi-purpose room.

When the students of the sample had been tested with both instruments, coefficients of correlation were calculated to determine the degree of relationship that existed between the scores of the DAT Form L and of the WISC.

According to Senders, "correlation means the reduction of unpredictability" (Senders, 1958). In this validity study Pearson's product-moment coefficient of correlation, r, was computed. This coefficient of correlation was a measure that expressed the amount of variance left in our prediction of one variable when we took the other variable into account. It could be used only when both variables came from interval or ratio scales, and could be used most sensibly only when the two variables were linearly related.

The Pearson product-moment was the basic measure of linear correlation. There was the assumption that both variables had been measured in interval or ratio scales and that the two variables were linearly related (Senders, 1958). Interval scaling by both tests was assumed on the basis of their construction. Linearity was assumed after inspection of the two-way scatterplot of scores. (This scatterplot was the basis for Figure I.) The machine formula for computing the Pearson product-moment correlation below was used (Senders, 1958),

$$r = \frac{N\Sigma XY - (\Sigma X)(\Sigma Y)}{\sqrt{\left[N\Sigma X^2 - (\Sigma X)^2\right]\left[N\Sigma Y^2 - (\Sigma Y)^2\right]}}$$

In these computations raw scores of the DAT and WISC Full Scale I. Q.s were used. Correlations were computed

for the two tests as follows: DAT VR+NA and WISC Full Scale, then DAT VR+NA+AR and WISC Full Scale for the total group and subgroups by sex.

It was felt that using the Full Scale I. Q. would present a more global concept than the Verbal or Performance I. Q.

A statistical test was necessary in order to determine whether or not the first and second null hypotheses of the study should be accepted or rejected.

To test the significance of a correlation between a set of paired observations, the null hypothesis was assumed equal to zero, the coefficient one would obtain between random values on the two measures. The test of significance was the distribution of t which employs Fisher's Z transformation of r. The formula to obtain the observed value of t was (Ferguson, 1959)

$$t = r \sqrt{\frac{N - 2}{1 - r^2}}$$

A test of significance was then applied by using the distribution of t. The t score is a statistic which can show the exact degree to which the data differs from what could be expected as result of chance. Values of t are arranged in statistical tables from which precise interpretations of data are possible. The .01 level of confidence (or one percent level) used in this study means that the difference which yielded this t value as a result of chance only once in a hundred times.

> The number of degrees of freedom associated with this value of t is N - 2. The loss of two degrees of freedom results because testing the significance of the slope of a regression line from zero (Ferguson, 1959).

The number of degrees of freedon (df) associated with the variability about a straight line fitted to a set of points was two less than the number of observations. A straight line would always fit two points with no possible freedom to vary.

A statistical calculation to determine whether the third null hypothesis should be accepted or rejected was also necessary.

In order to test the significance of the difference between the scores of the local sample and the national norm sample, the difference between the means of the two groups was tested. To test two means where the population variances were unequal, it was necessary to have available the means and standard deviations for both the Full Scale I. Q. scores of the study group and Full Scale I. Q. scores of the publisher's norm group. The formula used for calculating the mean was (Cottle and Downie, 1960)

$$X = \frac{\Sigma X}{N}$$

and the formula used for calculating the standard deviation was (Cottle and Downie, 1960)

$$s = \sqrt{\frac{x^2}{N - 1}}$$

Cochran and Cox provide an adjustment in the t formula required when variance are unequal. By using their method, it was necessary first to calculate the standard error of the differences between the two means, using the following formula (Ferguson, 1959)

$$s_{\bar{x} - \bar{y}} = \sqrt{\frac{(X - \bar{X})^2}{N_x(N_x - 1)} + \frac{(Y - \bar{Y})^2}{N_y(N_y - 1)}} = \sqrt{s_{\bar{x}}^2 + s_{\bar{y}}^2}$$

The differences between the sample means was then divided by the standard error of the difference to obtain the usual t ratio where a difference is divided by the standard error of that difference (Ferguson, 1959)

$$t = \frac{\bar{X} - \bar{Y}}{s_{\bar{x} - \bar{y}}}$$

When one sample was based on N cases with N - 1 df and another sample was based on N cases with N - 1 df, it was necessary to obtain the critical value of t required for significance at the one percent level then interpolate the approximate value of t required for significance at the one percent level with the formula (Ferguson, 1959)

$$t_{.01} = \frac{s_x^2 t_x + s_y^2 t_y}{s_x^2 + s_y^2}$$

The value of t obtained by dividing the difference between means by the standard error of their difference had to be equal to or greater than $t_{.01}$ before significance at the one percent level could be claimed.

In this study the probability of .01 level of confidence as a standard of significance was used. This seemed adequately critical in evaluating the results of this type of study.

CHAPTER IV

RESULTS OF THE STUDY

The study concerning the concurrent validity of two testing instruments used for one hundred ninth grade students at Coronado Junior High School yielded much data, some of which would appear to have relevance for the local school district.

A tabulation of the coefficients of correlation will be presented, supplemented by general information about the results of each test.

I. CONCURRENT VALIDITY

DAT raw scores for VR+NA, Verbal Reasoning, Numerical Ability, Abstract Reasoning, and WISC Verbal, Performance, and Full Scale I. Q. scores were tabulated for the males and females in Appendix A, using a code number to identify each student.

Coefficients of correlation were calculated for the DAT VR+NA and WISC Full Scale, and for the DAT VR+NA+AR and WISC Full Scale for the total group and for male and female subgroups. These correlations were then tested to see if they were significantly better than chance. The calculations are summarized in Appendix B. The practical significance of these coefficients as indicators of concurrent validity is discussed in Chapter V.

Since the major purpose of this study was to show the relationship between the VR+NA and Full Scale I. Q., this comparison is discussed first and is considered the primary one.

An r of .66 was obtained between scores from the DAT Form L VR+NA and the WISC Full Scale. For the fifty boys in the sample the correlation was .55, and for the fifty girls the correlation was .79, using the VR+NA and the Full Scale.

The correlation for the VR+NA and the Full Scale for the total group appeared to be reasonably high, but not as high as that reported in the study between the VR+NA and WAIS Full Scale for the Federal Reformatory inmates (Doppelt and Seashore, 1959).

The correlation for the total group between the scores from the VR+NA and Full Scale was higher than the one obtained by the boys, but lower than the one calculated for the girls. There appeared to be more relationship for DAT VR+NA and WISC Full Scale for girls than for either the total group or the boys.

Because there were indications that the Verbal Reasoning and Numerical Ability tests might be educationally loaded, the Abstract Reasoning Test was added for further validation study.

Combining the scores of the boys and girls, an r of .66 was obtained for the three DAT tests--Verbal Reasoning, Numerical Ability and Abstract Reasoning--and the WISC Full

Scale. Note that this is identical to the one obtained when using just the VR+NA and Full Scale scores for the total group.

For the fifty boys, the correlation between the VR+NA+AR and the WISC Full Scale was .54; this correlation was similar to that computed for the DAT Verbal Reasoning and Numerical Ability Tests with the WISC Full Scale.

A coefficient of correlation of .78 was obtained for girls between the Verbal Reasoning, Numerical Ability, Abstract Reasoning Tests and the WISC Full Scale. This figure was very little different from the .79 figure for the VR+NA and Full Scale.

Testing the Null Hypotheses

The coefficient of correlation between the DAT VR+NA and WISC Full Scale for scores of all the students in the study was .66. The first null hypothesis stating there was no difference between these scores and scores with no correlation had to be rejected at the 1% level of significance. The null hypotheses were also rejected at the 1% level for the subgroups of males and females.

The second null hypothesis stating there was no difference between the DAT tests, Verbal Reasoning, Numerical Ability, Abstract Reasoning and WISC Full Scale, and scores with no correlation was rejected at the 1% level for the young people in the study. In addition the null hypotheses applied to males and female groups considered separately

were rejected at the 1% level of significance for the VR+ NA+AR compared with WISC.

The third null hypothesis that there was no difference between the mean tested intelligence, as measured by the WISC Full Scale for the ninth grade students at Coronado Junior High School, and ninth grade students in WISC published norms (Wechsler, 1949) was accepted at the 1% level for the total group and male and female subgroups. Because the obtained value of t was less in this case than the critical value required, it was concluded that the means were not significantly different.

II. DAT DATA

In order to offer a comparison, the results concerning the boys and girls were treated together and then separately by subgroups; this seemed to be defensible because the class from which the study group was selected was equally divided between boys and girls.

Raw scores of the DAT tests which were pertinent and appropriate were listed. Since the DAT Manuals treat boys and girls' data separately, raw scores for the total group were tabulated, but they were not considered extensively.

Mean raw scores on the DAT tests for the hundred subjects of this study were: VR+NA, 43.6; Verbal Reasoning, 24.1; Numerical Ability, 19.6; and Abstract Reasoning, 33.4.

Mean raw score on the VR+NA for the fifty boys was 42.6 compared with a national mean of 43.3 (Bennett et al.,

1966). The girls who participated in the study on the VR+NA had a mean raw score of 44.6 compared to 42.5 for a national average for ninth grade girls (Bennett et al., 1966). Indications were the boys appeared to be a little below the national mean raw score for the VR+NA (index of scholastic aptitude), whereas the girls were a little above the mean.

The mean raw score on Verbal Reasoning for the fifty boys was 23.4 while the national mean raw score for ninth grade boys was 22.4 (Bennett et al., 1966). For the girls on the same test their mean raw score was 25.0 compared with a national mean raw score for ninth grade girls of 21.9 (Bennett et al., 1966).

The fifty boys had a mean raw score of 19.6 on the Numerical Ability Test, whereas the fifty girls had a mean raw score of 19.5. The national mean raw scores on this test were 20.9 and 20.6 respectively (Bennett et al., 1966).

On Abstract Reasoning the fifty boys had a mean raw score of 33.6 compared to the national mean for this grade level which was 29.3 (Bennett et al., 1966). The girls in this study had a mean score of 33.1 compared to the national average of 28.7 for the same test (Bennett et al., 1966).

III. WISC DATA

The students in the sample had a mean WISC Full Scale I. Q. of 100.8, which is in the average range according to the WISC Manual (Wechsler, 1949). Sixty-one percent of them had Full Scale I. Q.s between 90 and 110, which is defined

as the average range by the Manual. Twelve of the students had Full Scale I. Q.s above a plus one standard deviation, and eight of them had Full Scale I. Q.s below a minus one standard deviation.

A mean Full Scale I. Q. of 100.5 was obtained by the boys. In a range of Full Scale I. Q.s from 80 to 130 listed for boys in the tables, sixty percent are spread between 90 and 110.

The girls' Full Scale I. Q. ranged from 81 to 127; there were sixty-two percent of them who had Full Scale I. Q.s between 90 and 110. The mean Full Scale I. Q. for the fifty girls was 101.2, which differed little from that of the boys.

The mean Verbal I. Q. for the sample was 97.8, which would be considered average according to the WISC Manual.

Verbal I. Q. scores for boys ranged from 76 to 126; of the fifty boys tested sixty-four percent of them had average Verbal I. Q. Their mean Verbal I. Q. was 98.6, which would be considered average.

The girls' Verbal I. Q. scores ranged from 75 to 123; sixty-six percent of the girls had Verbal I. Q.s considered average. Their mean Verbal I. Q. was 96.6 which was within the average range, but a little lower than the mean Verbal I. Q. obtained by the boys.

The mean Performance I. Q. for the total group was 104.1; this was 6.3 I. Q. points higher than the mean Verbal I. Q. for the boys and girls who participated in the study.

A mean Performance I. Q. of 102.4 was obtained by the boys. The range was from 79 to 128; forty-six percent of the boys had average Performance I. Q., which was less than half the males tested.

The girls' Performance I. Q. scores ranged from 82 to 131 with a total mean Performance I. Q. of 105.8. Twenty-five girls, representing fifty percent of the girls who participated in the study, had average Performance I. Q.s. The girls' mean Performance I. Q. was a little higher than that obtained by the boys.

Expectancy tables are presented in Figures I, II and III on the following pages for the total group and the male and female subgroups. DAT VR+NA raw scores will be given horizontally and WISC Full Scale I. Q. scores will be listed vertically.

These data were presented to indicate graphically how the scores made by ninth grade students who were participants in the validation study on the two instruments concurred.

The numbers in the body of each expectancy table are percents of the row total shown at the extreme right of the table. For example, the two numbers in the top row on the expectancy table in Figure I would be interpreted as follows: only two students in the study scored on the WISC Full Scale I. Q. between 107-132; one of these (or fifty percent of the two students) scored between 52-58 on the DAT VR+NA; the other student scored between 73-79. Interpreting the figures in

the first column: fourteen percent of the fourteen students who scored between 86-91 on the WISC Full Scale scored between 14-20 on the DAT VR+NA.

FIGURE I

EXPECTANCY TABLE FOR TOTAL GROUP

WISC Full Scale I. Q.	14-20	21-27	28-34	35-41	42-48	49-51	52-58	59-65	66-72	73-79	Total n
107-132							50			50	2
121-126							20	60		20	5
115-120			12		12	12	25	25	12		8
109-114		7	14		14	21	21	14		7	14
103-108			15	15	8		46	8	8		13
98-102		7		33	13	13	7	20	7		15
92- 97		11	28	33	22		6				18
86- 91	14	21	21	14	7	7	14				14
80- 85	18	27	45	9							11
Total n	4	10	18	16	11	7	17	11	3	3	100

DAT VR+NA

FIGURE II

EXPECTANCY TABLE FOR MALES

WISC Full Scale I. Q.	14-20	21-27	28-34	35-41	42-48	49-51	52-58	59-65	66-72	73-79	Total n
127-132							100				1
121-126							33	33		33	3
115-120			25			25	25	25			4
109-114		17	17			17	33	17			6
103-108			20	20			40		20		5
98-102		14		29	14			29	14		7
92- 97		11	22	33	22		11				9
86- 91	22	11	11	11	11	11	22				9
80- 85	16	16	67								6
Total n	3	5	10	7	4	3	10	5	2	1	50

DAT VR+NA

FIGURE III

EXPECTANCY TABLE FOR FEMALES

WISC Full Scale I. Q.	14-20	21-27	28-34	35-41	42-48	49-51	52-58	59-65	66-72	73-79	Total n
127-132										100	1
121-126								100			2
115-120					25		25	25	25		4
109-114			12		25	25	12	12		12	8
103-108			12	12	12		50	12			8
98-102				38	12	25	12	12			8
92- 97		11	33	33	22						9
86- 91		40	40	20							5
80- 85	20	40	20	20							5
Total n	1	5	8	9	7	4	7	6	1	2	50

DAT VR+NA

CHAPTER V

CONCLUSIONS AND RECOMMENDATIONS

Several conclusions and recommendations are presented below as a result of the concurrent validity study of one hundred ninth grade students at Coronado Junior High School.

First, conclusions of the statistical testing of the hypotheses will be stated.

I. TEST OF HYPOTHESES

The first null hypothesis that there was no difference between the WISC Full Scale and the DAT VR+NA score for the students in the study was rejected at the 1% level for the total group and for the boys and girls subgroups. Thus, it is concluded that a difference does exist between the scores of the two test instruments.

By adding the Abstract Reasoning Test to the Verbal Reasoning and the Numerical Ability Tests of the DAT did not contribute to the coefficients of correlation with the WISC Full Scale for the students in the sample. Consequently, the second null hypothesis that there was no difference between the WISC Full Scale score and the DAT VR+NA+AR scores for the students participating in the study was likewise rejected at the 1% level of significance for the total group and male and female subgroups. Thus, it is concluded that a difference does exist between the scores of the two tests.

Cottle and Downie (1960) in their book Procedures and Preparation for Counseling presented an index for forecasting efficiency of coefficients of correlation of various sizes:

r	Index of Forecasting Efficiency (%)
.80	40
.70	28.6
.65	24
.60	20
.55	16.5
.50	13.4
.45	10.7
.40	8.3
.35	6.3
.30	4.6

This information helps one judge whether the correlations obtained indicate relationships which are of practical significance.

The relationship of the WISC Full Scale and DAT VR+NA for the sample appeared to be able to predict roughly 24% better than chance. Using the Full Scale and VR+NA for girls, the prediction can be made close to 40% better than chance, whereas for boys the index for forecasting efficiency dropped to 16.5%. The correlations of .66 for both sexes, .55 for boys and .79 for girls did appear to be sufficiently high to warrant not giving an individual intelligence test. On the basis of this information the author concludes that it is unnecessary to give an individual intelligence when the results of a scholastic aptitude test are available.

Scores of the group indicate that approximately two out of three boys who participated in the study had average

Verbal intelligence; about one-half of them had average Performance intelligence. However, six out of ten participants on the WISC Full Scale rated average.

The variation of scores for girls tested revealed two-thirds had average Verbal intelligence but one-half of them had average Performance intelligence. On the WISC Full Scale six out of ten rated average intelligence.

Even though a majority of the boys and girls in the sample were categorically considered average, one out of three on the Verbal Scale, one out of two on the Performance Scale, and four out of ten on the Full Scale fell outside the range that was considered average according to the WISC Manual (Wechsler, 1949). Some of the students might have had average verbal intelligence, and above average performance intelligence, or any other similar combinations.

The boys participating in the study had a higher mean Verbal I. Q. than the girls; the average for the boys was 98.6 whereas for the girls it was 96.7. The reverse was indicated in the case of the mean Performance I. Q.: the boys' mean Performance I. Q. was 102.4 compared to 105.8 for the girls.

The boys and girls from Coronado had a higher mean raw score on the Verbal Reasoning of the DAT than the national mean. It was apparent from their scores that the students were able to think constructively in words.

Neither the fifty boys as a whole nor the fifty girls as a whole scored as high as the national mean on the DAT

Numerical Ability Test. In the eighth grade the students from Coronado had modern math, whereas in the ninth grage, some of the students had traditional general math while others had the new approach to algebra. Perhaps this may have affected their scores on the Numerical Ability Test which was designed to test arithmetic computation (Bennett et al., 1966).

The students did better as a whole than the national mean on the Abstract Reasoning which would indicate they were able to reason non-verbally.

The results of the study indicated that the mean Full Scale I. Q. for the students was average: sixty-one percent of them fell in this range. It could be generalized that probably six out of ten ninth graders at Coronado Junior High School had average tested intelligence, since those who participated in the study were chosen at random.

Due to the apparent similarities of the other junior high school in the District to that of Coronado, probably six out of ten of their ninth graders had average tested intelligence, also. This might be further generalized to be applicable to the seventh and eighth grade boys and girls at Coronado. It would seem any assertion that Coronado Junior High School had a large segment of dullards would seem to be unfounded and unwarranted. It is therefore concluded that there is no difference in tested intelligence as measured with the WISC Full Scale I. Q. between ninth grade

students of Coronado Junior High School and ninth grade students in WISC published norms.

In all probability the results of the data from the study could represent a junior high school in a middle-class suburban location of a metropolitan area in the Middle West.

II. RECOMMENDATIONS

Several topics for further study are suggested below. From the testing and reading that had been done, some observations that might warrant further investigation and research will be included.

First, the fact that the boys in the study had a higher mean WISC Verbal I. Q. and lower mean Performance I. Q. than the girls might indicate an area for additional research and study.

The underlying reasons why the WISC Full Scale and DAT VR+NA had a considerably higher coefficient of correlation for ninth grade girls than for ninth grade boys at Coronado Junior High School might be worth further inquiry.

A similar validation study using the WISC Full Scale and DAT Verbal Reasoning, Numerical Ability and Abstract Reasoning might prove enlightening; such a study might substantiate or refute what had been found in this study.

What effect that the new approach to math had on the Numerical Ability Test should warrant some investigation. The DAT Manual (fourth edition) listed several pages of

validity studies, but there was no indication whether or not these were influenced by the new approach to math.

APPENDIXES

APPENDIX A

DAT AND WISC RAW DATA

APPENDIX A

DAT AND WISC RAW DATA FOR BOYS

	WISC			DAT			
Boys	Verb.	Perf.	Total	VR+NA	VR	NA	AR
1	96	106	101	67	38	29	46
2	106	97	102	59	31	28	26
3	91	113	101	24	14	10	35
4	86	78	80	31	21	10	39
5	97	124	111	64	32	32	43
6	111	133	124	75	41	34	47
7	92	111	101	35	21	14	35
8	97	85	91	32	25	7	31
9	76	90	81	27	12	15	34
10	111	111	112	57	33	24	37
11	106	110	109	53	28	25	32
12	121	120	123	57	38	19	44
13	90	101	95	26	9	17	31
14	108	101	105	29	13	16	27
15	101	96	99	62	33	29	43
16	100	87	93	45	23	22	34
17	91	106	98	40	17	23	40
18	116	120	120	52	33	19	39
19	97	115	107	40	15	25	16
20	125	117	123	59	32	27	41
21	108	113	111	49	23	26	40
22	86	94	89	51	29	22	36
23	100	92	96	58	35	23	43
24	92	79	85	29	20	9	21
25	94	90	91	47	23	24	38
26	85	111	97	35	17	18	40
27	103	80	91	54	33	21	36
28	91	80	85	14	10	4	12
29	81	104	91	20	8	12	31
30	108	113	111	29	23	6	26
31	92	100	96	28	12	16	32
32	86	93	88	21	10	11	13
33	84	85	83	34	20	14	34
34	89	96	91	17	6	11	17
35	100	80	90	52	24	28	31
36	99	120	109	25	15	10	28
37	109	96	103	68	33	35	37
38	103	101	102	47	30	17	29
39	110	100	106	55	32	23	33
40	113	115	115	48	46	22	33

41	91	96	93	28	12	16	19
42	111	120	117	34	20	14	37
43	106	121	115	59	30	29	44
44	85	104	93	35	15	20	40
45	91	82	85	32	13	19	32
46	89	96	91	35	20	15	21
47	96	90	93	43	19	24	34
48	92	106	99	37	20	17	40
49	95	113	104	56	29	27	43
50	126	128	130	58	36	22	41

DAT AND WISC RAW DATA FOR GIRLS

	WISC			DAT			
Girls	Verb.	Perf.	Total	VR+NA	VR	NA	AR
1	94	122	108	52	24	28	37
2	96	100	98	39	17	22	33
3	99	108	104	29	13	16	22
4	109	131	121	59	36	23	45
5	91	114	102	52	31	21	40
6	123	127	127	78	46	32	42
7	82	113	96	26	18	8	22
8	95	108	101	43	22	21	36
9	86	89	86	33	20	13	30
10	96	110	103	45	29	16	26
11	108	108	109	51	29	22	29
12	113	122	119	62	42	20	46
13	91	106	98	40	23	17	34
14	97	113	105	54	31	23	39
15	82	87	83	34	21	13	38
16	92	94	93	47	24	23	40
17	110	86	99	49	31	18	35
18	104	103	104	38	26	12	29
19	81	82	80	36	16	20	28
20	92	96	93	38	16	22	22
21	94	93	93	33	21	12	20
22	111	117	115	48	29	19	39
23	89	92	89	26	16	10	22
24	104	117	111	46	24	22	39
25	99	94	96	39	18	21	36
26	99	121	110	31	18	13	26
27	91	117	104	54	31	23	40
28	101	115	109	44	23	21	30
29	106	124	116	58	33	25	43
30	116	107	113	61	27	34	37
31	97	111	104	59	29	30	41
32	82	107	93	31	14	17	24
33	82	89	84	26	12	14	32
34	79	99	87	31	15	16	21
35	101	111	107	55	23	32	45
36	90	110	99	36	21	15	36
37	90	90	89	27	21	6	32
38	84	101	91	39	27	12	32
39	81	83	80	22	14	8	8
40	103	115	109	52	29	23	35
41	120	118	121	60	39	21	44
42	110	108	110	78	42	36	42
43	103	118	111	50	28	22	36
44	75	92	81	19	8	11	13
45	99	96	97	32	14	18	27

46	95	107	101	61	31	30	38
47	113	122	119	72	45	27	40
48	92	100	96	47	35	12	27
49	103	100	101	49	31	18	40
50	97	97	97	41	20	21	38

APPENDIX B

TESTING SIGNIFICANCE

APPENDIX B

TESTING SIGNIFICANCE

DAT VR+NA AND WISC FULL SCALE (TOTAL GROUP)

$$H_1 = r_{obs.} - r_0 = 0$$

$$\alpha = t_{.01}^{98} = 2.660 \text{ (critical value) (df = N - 2)}$$

$$t_{obs.} = r\sqrt{\frac{N - 2}{1 - r^2}}$$

$$r = \frac{N\Sigma XY - (\Sigma X)(\Sigma Y)}{\sqrt{\left[N\Sigma X^2 (\Sigma X)^2\right]\left[N\Sigma Y^2 - (\Sigma Y)^2\right]}}$$

$$r = \frac{100\ (451488) - (1088)\ (4364)}{\sqrt{\left[100\ (1031484) - (10088)^2\right]\left[100\ (211144) - (4364)^2\right]}}$$

$$r = .66$$

$$t_{obs.} = .66\sqrt{\frac{100 - 2}{1 - (.66)^2}}$$

$$t_{obs.} = 8.7318$$

Conclude: Since $t_{obs.}$ is greater than the critical value of t , the null hypothesis is rejected and the conclusion is drawn that there is a significant correlation between the DAT VR+NA and WISC Full Scale.

TESTING SIGNIFICANCE

DAT VR+NA AND WISC FULL SCALE (BOYS)

$$H_1 = r_{obs.} - r_0 = 0$$

$$\alpha = t^{48}_{.01} = 2.704 \text{ (critical value) (df = N - 2)}$$

$$t_{obs.} = r\sqrt{\frac{N - 2}{1 - r^2}}$$

$$r = \frac{N\Sigma XY - (\Sigma X)(\Sigma Y)}{\sqrt{\left[N\Sigma X^2 (\Sigma X)^2\right]\left[N\Sigma Y^2 - (\Sigma Y)^2\right]}}$$

$$r = \frac{50\,(219271) - (5026)(2132)}{\sqrt{\left[50\,(512532) - (5026)^2\right]\left[50(102052) - (2132)^2\right]}}$$

$$r = .55$$

$$t_{obs.} = .55\sqrt{\frac{50 - 2}{1 - (.55)^2}}$$

$$t_{obs.} = 4.5540$$

Conclude: Since $t_{obs.}$ is greater than the critical value of t , the null hypothesis is rejected and the conclusion is drawn that there is a significant correlation between the DAT VR+NA and WISC Full Scale.

TESTING SIGNIFICANCE

DAT VR+NA AND WISC FULL SCALE (GIRLS)

$$H_1 = r_{obs.} - r_0 = 0$$

$$\alpha = t^{48}_{.01} = 2.704 \text{ (critical value) (df = N - 2)}$$

$$t_{obs.} = r\sqrt{\frac{N - 2}{1 - r^2}}$$

$$r = \frac{N\Sigma XY - (\Sigma X)(\Sigma Y)}{\sqrt{[N\Sigma X^2 (\Sigma X)^2][N\Sigma Y^2 - (\Sigma Y)^2]}}$$

$$r = \frac{50\ (232177)\ (5062)\ (2232)}{\sqrt{[50\ (518952) - (5062)^2][50\ (109092) - (2232)^2]}}$$

$$r = .79$$

$$t_{obs.} = .79\sqrt{\frac{50 - 2}{1 - (.79)^2}}$$

$$t_{obs.} = 8.8796$$

Conclude: Since $t_{obs.}$ is greater than the critical value of t, the null hypothesis is rejected and the conclusion is drawn that there is a significant correlation between the DAT VR+NA and WISC Full Scale.

TESTING SIGNIFICANCE

DAT VR+NA+AR AND WISC FULL SCALE (TOTAL GROUP)

$$H_2 = r_{obs.} - r_0 = 0$$

$$\alpha = t_{.01}^{98} = 2.660 \text{ (critical value) (df = N - 2)}$$

$$t_{obs.} = r\sqrt{\frac{N - 2}{1 - r^2}}$$

$$r = \frac{N\Sigma XY - (\Sigma X)(\Sigma Y)}{\sqrt{\left[N\Sigma X^2 (\Sigma X)^2\right]\left[N\Sigma Y^2 - (\Sigma Y)^2\right]}}$$

$$r = \frac{100\,(793294) - (10088)(7701)}{\left[100\,(1031484) - (10088)^2\right]\left[100\,(638377) - (7701)^2\right]}$$

$$r = .66$$

$$t_{obs.} = .66\sqrt{\frac{100 - 2}{1 - (.66)^2}}$$

$$t_{obs.} = 8.7318$$

Conclude: Since $t_{obs.}$ is greater than the critical value of t , the null hypothesis is rejected and the conclusion is drawn that there is a significant correlation between the DAT VR+NA+AR and WISC Full Scale.

TESTING SIGNIFICANCE

DAT VR+NA+AR AND WISC FULL SCALE (BOYS)

$$H_2 = r_{obs.} - r_0 = 0$$

$$\alpha = t^{48}_{.01} = 2.704 \text{ (critical value) (df = N - 2)}$$

$$t_{obs.} = r\sqrt{\frac{N - 2}{1 - r^2}}$$

$$r = \frac{N\Sigma XY - (\Sigma X)(\Sigma Y)}{\sqrt{\left[N\Sigma X^2 (\Sigma X)^2\right]\left[N\Sigma Y^2 - (\Sigma Y)^2\right]}}$$

$$r = \frac{50\ (390351)\ (5026)\ (3813)}{\sqrt{\left[50\ (512532) - (5026)^2\right]\left[50\ (313987) - (3813)^2\right]}}$$

$$r = .54$$

$$t_{obs.} = .54\sqrt{\frac{50 - 2}{1 - (.54)^2}}$$

$$t_{obs.} = 4.4388$$

Conclude: Since $t_{obs.}$ is greater than the critical value of t , the null hypothesis is rejected and the conclusion is drawn that there is a significant correlation between the DAT VR+NA+AR and WISC Full Scale.

TESTING SIGNIFICANCE

DAT VR+NA+AR AND WISC FULL SCALE (GIRLS)

$$H_2 = r_{obs.} - r_0 = 0$$

$$\alpha = t^{48}_{.01} = 2.704 \text{ (critical value) (df = N - 2)}$$

$$t_{obs.} = r\sqrt{\frac{N - 2}{1 - r^2}}$$

$$r = \frac{N\Sigma XY - (\Sigma X)(\Sigma Y)}{\left[N\Sigma X^2 (\Sigma X)^2\right]\left[N\Sigma Y^2 - (\Sigma Y)^2\right]}$$

$$r = \frac{50\ (402943)\ (5062)\ (3888)}{\left[50\ (518952) - (5062)^2\right]\left[50\ (324390) - (3888)^2\right]}$$

$$r = .78$$

$$t_{obs.} = .78\sqrt{\frac{50 - 2}{1 - (.78)^2}}$$

$$t_{obs.} = 8.5878$$

Conclude: Since $t_{obs.}$ is greater than the critical value of t , the null hypothesis is rejected and the conclusion is drawn that there is a significant correlation between the DAT VR+NA+AR and WISC Full Scale.

TESTING SIGNIFICANCE

WISC FULL SCALE (TOTAL GROUP) AND FULL SCALE (PUBLISHER'S NORMS)

$$t_{.01} = \frac{s_{\bar{x}}^2 t_x + s_{\bar{y}}^2 t_y}{s_x^2 + s_y^2}$$

$$t_{.01} = \frac{(1.39)\ (2.660) + (1.13)\ (2.576)}{1.39 + 1.13}$$

$$t_{.01}^{98} = 2.63 \text{ (critical value) (df = N - 2)}$$

$$t_{obs.} = \frac{\bar{X} - \bar{Y}}{s_{\bar{x}} - \bar{y}}$$

$$t_{obs.} = \frac{100.8 - 100}{1.59}$$

$$t_{obs.} = .50$$

Conclude: Since $t_{obs.}$ is less than the critical value of t , the null hypothesis is accepted and the conclusion is drawn that there is no difference between the WISC Full Scale (Total Group) and Full Scale (Publisher's Norms).

TESTING SIGNIFICANCE

WISC FULL SCALE (BOYS) AND FULL SCALE (PUBLISHER'S NORMS)

$$t_{.01} = \frac{s_{\bar{x}}^2 t_x + s_{\bar{y}}^2 t_y}{s_x^2 + s_y^2}$$

$$t_{.01} = \frac{(2.98)\ (2.704) + (1.13)\ (2.576)}{2.98 + 1.13}$$

$$t_{.01}^{48} = 2.68 \text{ (critical value) (df = N - 2)}$$

$$t_{obs.} = \frac{\bar{X} - \bar{Y}}{s_{\bar{x} - \bar{y}}}$$

$$t_{obs.} = \frac{100.5 - 100}{2.03}$$

$$t_{obs.} = .25$$

Conclude: Since $t_{obs.}$ is less than the critical value of t , the null hypothesis is accepted and the conclusion is drawn that there is no difference between the WISC Full Scale (Boys) and Full Scale (Publisher's Norms).

TESTING SIGNIFICANCE

WISC FULL SCALE (GIRLS) AND FULL SCALE (PUBLISHER'S NORMS)

$$t_{.01} = \frac{s_{\bar{x}}^2 t_x + s_{\bar{y}}^2 t_y}{s_{\bar{x}}^2 + s_{\bar{y}}^2}$$

$$t_{.01} = \frac{(2.65)\ (2.704) + (1.13)\ (2.576)}{2.65 + 1.13}$$

$$t_{.01}^{48} = 2.66 \text{ (critical value) (df = N - 2)}$$

$$t_{obs.} = \frac{\bar{X} - \bar{Y}}{s_{\bar{x}} - \bar{y}}$$

$$t_{obs.} = \frac{101.2 - 100}{1.94}$$

$$t_{obs.} = .62$$

Conclude: Since $t_{obs.}$ is less than the critical value of t , the null hypothesis is accepted and the conclusion is drawn that there is no difference between the WISC Full Scale (Girls) and Full Scale (Publisher's Norms).

BIBLIOGRAPHY

BIBLIOGRAPHY

A. BOOKS

Boynton, Paul L. Intelligence, its manifestations and measurement. New York: D. Appleton, 1933.

Bennett, George K., Harold G. Seashore and Alexander G. Wesman. Differential Aptitude Tests manual. first edition. New York: Psychological Corporation, 1947.

_____. Differential Aptitude Tests manual. third edition. New York: Psychological Corporation, 1959.

_____. Differential Aptitude Tests manual. fourth edition. New York: Psychological Corporation, 1966.

Cottle, William C. and N. M. Downie. Procedures and preparation for counseling. Englewood Cliffs, New Jersey: Prentice-Hall, 1960.

Drever, James. A dictionary of psychology. Baltimore: Penquin Books, 1964.

Ferguson, George A. Statistical analysis in psychology and education. New York: McGraw-Hill, 1959.

Hall, Calvin S. and Gardner Lindzey. Theories of personality. New York: John Wiley and Sons, 1957.

Harriman, Phillip L. Handbook of psychological terms. Paterson, New Jersey: Littlefield, Adams, 1961.

Jordan, A. M. Measurement in education. New York: McGraw-Hill, 1953.

Revised standard version of Bible. New York: Thomas Nelson and Sons, 1952.

Senders, Virginia L. Measurement and statistics. New York: Oxford University Press, 1958.

Thorndike, Edward et al. The measurement of intelligence. New York: Bureau of Publications Teachers College, 1927.

Thorndike, Robert L. and Elizabeth Hagan. Measurement and evaluation in psychology and education. New York: John Wiley and Sons, 1961.

Wechsler, David. The measurement and appraisal of adult intelligence. Baltimore: Williams and Wilkins, 1958.

_____. Wechsler intelligence scale for children. New York: Psychological Corporation, 1949.

B. PUBLICATIONS OF TEST PUBLISHERS

Directions for administration and scoring and norms Forms L and M. New York: Psychological Corporation, 1963.

Educational Testing Service. Selecting an achievement test. Evaluation and Advisory Service Series No. 3. Princeton, New Jersey: Educational Testing Service, 1961.

Psychological corporation test catalog. New York: Psychological Corporation, 1965.

Test service bulletin, No. 38. New York: Psychological Corporation, 1949.

C. PERIODICALS

Altus, Grace Thompson. A note on the validity of the Wechsler Intelligence Scale for Children. J. Consult. Psychol., 1952, 16, 231.

Arnold, Frank C. and Winifred K. Wagner. Comparison of Wechsler Scale and Stanford-Binet scores for eight- and nine-year olds. J. Exp. Ed., 1955, 24, 91-94.

Bennett, George K., Harold G. Seashore and Alexander G. Wesman. The Differential Aptitude Tests: some comments by the authors. Occupations, 1948, 27, 20-22.

Burks, Harold F. and Paul Bruce. The characteristics of poor and good readers as disclosed by the Wechsler Intelligence Scale for Children. J. Ed. Psychol., 1955, 46, 488-93.

Chauncey, Henry. The rising trend of early aptitude testing. Nat. Assn. Sec. Sch. Prin. Bul., 1959, 43, 304-7.

Cottle, William C. The Differential Aptitude Tests: some comments. Occupations, 1948, 26, 344-45.

Delattre, Lois and David Cole. Comparison of the WISC and Wechsler-Bellevue. J. Consult. Psychol., 1952, 16 228-30.

Diamond, Solomon. The Wechsler-Bellevue Intelligence Scales and certain vocational aptitude tests. J. Psychol., 1947, 24, 279-82.

Doppelt, Jerome E. and Harold G. Seashore. Psychological testing in correctional institutions. J. Counsel. Psychol., 1959, 6, 81-92.

Dyer, Henry S. A psychometrician views human ability. Teach. Col. Rec., 1960, 61, 394-403.

Estes, Betty Worth. Influence of socio-economic status on Wechsler Intelligence Scale for Children: an exploratory study. J. Ed. Psychol., 1953, 17, 58-62.

Flanagan, John C. New developments in aptitude testing. Nat. Assn. Sec. Sch. Prin. Bul., 1955, 39, 50.

_____. Testing programs: their function in education. J. Sec. Ed., 1960, 35, 41-45.

Frandsen, Arden N. and Jay B. Higginson. The Stanford-Binet and the Wechsler Intelligence Scale for Children, J. Consult. Psychol., 1951, 15, 236-38.

Harlow, Justin et al. Preliminary study of the comparison between the Wechsler Intelligence Scale for Children and Form L of the Revised Stanford-Binet. J. Clin. Psychol., 1957, 13, 72-73.

Hills, John R. What I look for in a review of a scholastic aptitude test or battery. Personnel & Guid. J., 1964, 42, 711-14.

Holland, Glen A. A comparison of the WISC and Stanford-Binet iq's of normal children. J. Consult. Psychol., 1953, 17, 147-52.

Hunt, William et al. The validity of some abbreviated individual intelligence scales. J. Consult. Psychol., 1948, 12, 48-52.

Kureth, Sister Genevieve, Jean P. Muhr and Charles A. Weisgerber, S.J. Some data on the validity of the Wechsler Intelligence Scale for Children. Child Develop., 1952, 17, 281-87.

Krugman, Judith I., et al. Pupil functioning on the Stanford-Binet and the Wechsler Intelligence Scale for Children. J. Consult. Psychol., 1951, 15, 475-83.

Layton, William L. and Edward O. Swanson. Relationship of ninth grade Differential Aptitude Test scores to eleventh grade test scores and high school rank. J. Ed. Psychol., 1958, 49, 153-55.

Littell, William M. The Wechsler Intelligence Scale for Children: review of a decade of research. Psychol. Bul., 1960, 57, 132-56.

Lundy, Charles T. and Bruce Shertzer. Making test data useful. Personnel & Guid. J., 1963, 42, 62-3.

Malnig, Lawrence R. Anxiety and academic prediction. J. Counsel. Psychol., 1964, 11, 72-75.

Millman, Jason and Marvin D. Glock. Trends in the measurement of general mental ability. R. Ed. Res., 1965, 35, 17-24.

Mussen, Paul, Sanford Dean and Margery Rosenberg. Some further evidence on the validity of the WISC. J. Consult. Psychol., 1952, 16, 410-11.

Pastovic, John J. and George M. Guthrie. Some evidence on the validity of the WISC. J. Consult. Psychol., 1951, 15, 385-86.

Seashore, Harold G. Difference between verbal and performance iqs on the Wechsler Intelligence Scale for Children. J. Consult. Psychol., 1951, 15, 62-67.

Seashore, Harold, Alexander Wesman and Jerome Doppelt. The standardization of the Wechsler Intelligence Scale for Children. J. Consult. Psychol., 1950, 14, 99-110.

Segel, David. The validity of a multiple aptitude test at secondary school level. Ed. & Psychol. M., 1947, 7, 695-705.

Siegman, Aron W. The effect of manifest anxiety on a concept formation task, a nondirected learning task, and on timed and untimed intelligence tests. J. Consult. Psychol., 1956, 20, 176-78.

Stanley, Julian C. Developments and applications of tests of general mental ability. R. Ed. Res., 1953, 23, 11-32.

Stroud, James B., Paul Blommers and Margaret Lauber. Correlation of WISC and achievement tests. J. Ed. Psychol., 1957, 48, 18-26.

Wallen, Norman E. Development and application of tests of general mental ability. R. Ed. Res., 1962, 32, 15-24.

Warrington, Willard G. and Joe L. Saupe. Development and applications of tests of general mental ability. R. Ed. Res., 1959, 29, 15-25.

Weider, Arthur et al. The Wechsler Intelligence Scale for Children and the Revised Stanford-Binet. J. Consult. Psychol., 15, 330-33.

Williams, Nancy. A study of the validity of the Verbal Reasoning subtest and the Abstract Reasoning subtest of the Differential Aptitude Tests. Ed. & Psychol. M., 1952, 12, 129-31.

D. ENCYCLOPEDIA ARTICLES

Neidt, Charles O. Tests and measurement. World Book, 1966, 18, 142-44.