A NEW LOOK AT THE WONDERLIC PERSONNEL TEST

A Thesis

Presented to

the Faculty of the Department of Psychology

University of Houston

In Partial Fulfillment

of the Requirements for the Degree

Master of Arts

By

David De Fee

January, 1970

ACKNOWLEDGMENTS

I would like to express my fullest appreciation to the following college instructors and professors who made it possible to obtain subjects for this research study by generously donating valuable class time period(s) for testing purposes during the summer six-week sessions: Mrs. Norma Brady, Sam Houston University, Department of English; Dr. Gladys Ford, Texas Southern University, Department of English; Mrs. Amy Sarvis, Mr. Victor Vitanza, and Dr. William C. Wright, Department of English, University of Houston, and Mr. Mark Moore, University of Houston, Department of Psychology. I also thank Mr. Curtis Mabry, University of Houston, Department of Psychology, for donating his time and help in testing subjects at Texas Southern University.

In addition, I would like to extend a special acknowledgment of thanks to Dr. John A. Cox, University of Houston, Counseling and Testing Center, for his undivided time, interest, and enthusiasm in working with me on complex problems that arose during the formulation of some of the research designs and the statistical treatment of the research data.

TABLE OF CONTENTS

LIST OF TABLES

TABLE PAGE

A NEW LOOK AT THE WONDERLIC PERSONNEL TEST

An Abstract of a Thesis

Presented to

the Faculty of the Department of Psychology

University of Houston

---

in Partial Fulfillment

of the Requirements for the Degree

Master of Arts

By

David De Fee

January, 1970

ABSTRACT

The purpose of this study was to scrutinize a very popular and widely used mental ability or general intelligence test, the Wonderlic Personnel Test, in a very rigorous fashion in relation to three general questions: (a) Does the Wonderlic Personnel Test favor one ethnic group over another group to any great extent? (b) If the Wonderlic Personnel Test does in fact favor one ethnic group over another ethnic group, what might be the basis for this differentiation? (c) Are the claims made by the test author of the Wonderlic Personnel Test in regards to its internal characteristics valid?

In order to answer the first question, a sample of White college students (N=35) and a sample of Negro college students (N=27) were administered Wonderlic Personnel Test Form I under both the standard administration time limit of 12 minutes and under an unlimited (power) administration procedure. Each group was composed of summer college students of both sexes who were under 25 years of age and who had obtained 30-90 semester college hours. A comparison of the mean score for each group indicated that the Negro group scored considerably lower (significant at the .01 level) on Wonderlic Personnel Test Form I than the White group under both the standard time limit administration and under the power administration. The difference in mean gain obtained between the two groups under the power administration was not significant, however, and thereby

supported the hypothesis that Negroes will not benefit more than Whites on Wonderlic Personnel Test Form I when given additional testing time. Clearly, then the answer to the first question is: Yes, the Wonderlic Personnel Test (Form I) does favor Whites over Negroes.

The second question was, "If the Wonderlic Personnel Test does favor one ethnic group over another, what might be the basis for this phenomenon"? To answer this question mean scores were computed for both the Whites and the Negroes based upon test items whose content was verbal in nature. The result of this procedure showed that the Negroes had a significantly lower (at the .01 level) mean score than did the Whites under both the 12-minute administration and under the power administration. This finding answered the above question and supports the hypothesis that Negroes will perform less well than Whites on the verbal items contained within Wonderlic Personnel Test (Form I) under a standard administration time limit. It appears as though the verbal items within Wonderlic Personnel Test Form I constitute a major type of item content that distinguishes between the two ethnic groups.

In order to verify the internal characteristics claimed for the Wonderlic Personnel Test by its author, Wonderlic test forms I, II, and IV were administered to a sample of White summer school students (N=108). Mean interform correlation coefficients between paired test forms which took into account a noted practice effect ranged from +.68 to +.75, lower than the +.82 to +.94 range given in the Wonderlic Personnel Test Manual (1966). Correlation charts

between paired test forms indicated that there was a high degree of variability among subjects taking alternate Wonderlic test forms I, II, and IV. Due to the low interform correlation coefficients obtained and the high degree of variability among subjects taking these test forms, Wonderlic Personnel Test Forms I, II, and IV can not be regarded as equal, equivalent, or alternate test forms. And the statistical treatment of the data to check on per cent of difficulty for units of ten questions between test forms showed that Wonderlic Personnel Test Forms I, II, and IV are not equal in terms of per cent of difficulty for units of ten questions.

In summary, the results of this study are at odds with the information found in the Wonderlic Personnel Test Manual (1966) and cast doubts on the adequacy of the interform reliability of these tests. Replication of this study in additional samples to test the generalizability of estimates of interform reliability of the Wonderlic should be performed.

CHAPTER I

INTRODUCTION TO THE WONDERLIC PERSONNEL TESTS

The Wonderlic Personnel Tests are composed of seven test forms regarded by the author to be comparable and similar: A, B, I, II, IV, V, and EM. These test forms were designed for testing adults in business and industrial situations where decisions had to be made by personnel executives in choosing one applicant among many for a given job within an industrial organization--selection; and in choosing from a number of jobs within an industrial organization, the one job best suited to the strengths and weaknesses of a particular applicant--placement. (Guion, 1965).

Each test form is composed of fifty test items of both the completion and multiple-choice types, and is administered under a 12-minute time limit either to a single individual or jointly to a group of individuals. Specific items within each test form include analogies, analysis of geometric figures, arithmetic problems involving computation and reasoning, word definitions, disarranged sentences. formal syllogisms, direction following using clerical items, similarities, sentence parallelism with proverbs, spatial relations items, and "commonsense" reasoning items

The raw score--number of questions answered correctly--obtained by an examinee taking one of the Wonderlic test forms is regarded by the test author to be indicative of the examinee's "mental ability"

level, and more precisely is thought of as an indicator of the extent to which an examinee can "(a) understand and think in terms of words, (b) understand and think in terms of numbers, (c) think in terms of symbols, and (d) think in terms of ideas "(Manual, 1966, p. 4)."

The historical evolution of the Wonderlic Personnel Test forms is described in an article published by Wonderlic & Hovland (1939):

> The characteristics most valued in an intelligence test for business and industry are somewhat different from those stressed in academic use. For both purposes, of course, adequate reliability and validity are required. But simplicity of administration and ease and objectivity of scoring become additional important factors in large scale industrial employment. The Otis Self-Administering Test (Higher Form) has become widely used in business and industry because of the very fact that it possesses these latter characteristics. This test, however, standardized and validated upon school children, has been shown to have a number of weaknesses which make it unsatisfactory for adult groups (Hovland and Wonderlic (1)). The present paper describes an abridgement of the Otis test, standardized upon adults in business and industry, which remedies the principal defects of the Otis test and still retains its essential advantages. The abridgement, labelled the "Personnel Test," requires only twelve minutes, is self-administering, easy to score, reliable, and valid and useful in a number of industrial situations (p. 685).

Thus, the Wonderlic Personnel Tests are in effect an outgrowth of the Otis (Higher Form), and actually were constructed from parts of the original four forms of the Otis in terms of test items.

Using over 8,000 Otis S-A tests, Wonderlic and Hovland (1939) proceeded to determine the validity and difficulty of the individual items in the four forms of the Otis Test. Then items which proved valid according to a schema of three successive criteria were selected to be included within three parallel forms of the new "Personnel Test".

To be included within the "Personnel Test" an item had to prove valid according to these three successive criteria. (a) first the item had to differentiate between successful and unsuccessful industrial employees. This procedure was accomplished by forming two criterion groups: Group I was made up of employees given high ratings by supervisors and having a record of two or more salary raises and at least one promotion; Group II contained employees with poor supervisory ratings and an unsatisfactory work history record. Groups I and II were composed exclusively of employees and ex-employees of the Household Finance Corporation. Only Otis test items which successfully differentiated between these two groups in terms of percentages of correct responses were retained. (b) secondly, items were selected which differentiated between good and poor school records based on data available on 350 Northwestern University students who were equivalent in age and year in college. Only Otis test items which showed statistically significant differences between students in the upper and lower twenty-five percent with respect to academic achievement were retained. (c) thirdly, the final criterion for each test item was based on biserial coefficients of correlation between total score on the test and the passing or failing of the individual item. (Wonderlic and Hovland, 1939, p. 688).

Through this careful selection of items, Wonderlic and Hovland (1939) attempted to remedy three principal defects that they found within the Otis as used in industrial testing situations: (a) too many of the questions within the Otis were extremely easy for adults, mainly because the Otis was standardized on children, (b) because of the

excess number of easy questions contained within the Otis, too many examinees were completely finishing the test, and (c) the test items within the Otis were inadequately arranged in terms of their difficulty level so that many of the adults were able to obtain disproportionately high scores by skipping about and selecting the easy items.

Wonderlic & Hovland (1939) commented on how they attempted to correct these three defects in the following manner:

> 1. In the original Otis over a quarter of the questions were passed by between ninety and one hundred per cent of the individuals tested. Over sixty per cent of the items were passed by seventy-five or more per cent of the subjects. When items are passed by such a high percentage of individuals they provide little basis for differentiation of ability. . . . In constructing the Personnel Test care was taken to include questions representative of the entire range of difficulty, with an average such that about fifty per cent would pass and fifty per cent would fail the item of mean difficulty (pp. 686-687).
>
> 2. In the original Otis Test as many as forty per cent of some groups finished the test in the thirty-minute time limit. This, of course, results in under-measurement of the brighter individuals in the group. (In constructing the Personnel Test) the length of the test was made such that about two to five per cent of average groups complete the test in the twelve-minute time limit (p. 688).
>
> 3. The third difficulty with the Original Otis Test is that the items are very inadequately arranged in order of difficulty. The correlations between the difficulty of the questions and the printed order were only from +.42 to +.75 for the groups which we studied. This permits individuals to get disproportionately high scores by skipping about and selecting the easy items. In the Personnel Test extreme care was taken to have the items uniformly increase in difficulty. . . . The results of the restandardization are shown in correlations (p) of +.90 to +.95 (P.E.'s=.009-015) between the order in which the items are printed in the Personnel Test and their difficulty in terms of the percentage of sub-

> jects passing them. These data are based upon all three forms, and upon diverse groups of subjects. When corrected for attenuation the correlations are all +.95 or higher (pp. 689-690).

After these corrections were made and three new forms of the Personnel Test constructed, the authors felt that they had created three forms of a new test which were closely equated in terms of average difficulty and were thereby similar and comparable.

> In addition to being equated in average difficulty the forms are equivalent by successive fifths throughout the examination . . . . The percentages of subjects passing each individual question were first determined; from these the averages for successive groups of ten were computed (p. 693).

The authors identified these three comparable and parallel Personnel Test forms as Forms "D", "E", and "F".

Since 1939 Wonderlic Personnel Test forms D, E, and F have been replaced by newer forms, and none of these original forms are now available. Forms A and B were constructed in 1942 using the same criteria as used in the construction of the original forms. However, Forms A and B included a greater diversity of items in terms of content and type than were included in Forms D, E, and F. Likewise, additional forms--I, II, IV, V, and EM--were developed in 1959, with Form EM being restricted in terms of sale only to fee employment agencies.

At present only Forms A, B, I, II, IV, V, and EM are available to personnel executives for usage in the selection and placement process. However, all of these forms retain and commonly share these general features, according to the test Manual (1966), and thus can

be considered equivalent forms: (a) "The questions selected range in difficulty from extremely easy to extremely difficult for adult groups (p.3)," (b) "There is an equality in the steps of difficulty throughout the testing range (p.3)," (c) "There is a normal distribution of difficulties with a mean about 60% (p.3)," (d) "The correlations between position in test and item difficulty are all above .90 (p.3)," (e) "Comparing the results of one 12-minute test taken immediately after another gave correlations of .82 to .94, indicating good reliability (p. 3)," (f) "The forms A, B, I, II, IV, V, and EM have been so devised as to be equal in per cent of difficulty for units of ten questions, as well as for total score (p.4)," and (g) "We know that the questions get progressively more difficult (p. 4)."

CHAPTER II

REVIEW OF RELEVANT LITERATURE ON THE WONDERLIC PERSONNEL TESTS

Although there is an abundance of research literature reported on the Wonderlic Personnel Test forms in various business publications and professional journals in regards to quantitative norming for various occupational categories and groups, it is clear that there is a paucity of professional research studies which have investigated the validity of the internal characteristics of the test forms per se as reported as established fact by the test author. In this section a review has been made of those professional research studies which have attempted to support or refute the claims of the validity of the internal characteristics of the Wonderlic test forms as reported in Chapter I.

TEST ADMINISTRATION TIME LIMIT

Wright & Laing (1943) tested the statement made during a panel discussion of testing at a session of the American Management Association that "the time limit for the Wonderlic Personnel Test should be ignored inasmuch as the average individual was likely to be penalized by the time interval (p. 315)." Specifically, their hypothesis was that any extension of time beyond that set by the author of the Wonderlic (12 minutes) would not change the relative rank of an applicant within an industrial testing situation.

In order to test their hypothesis Wright & Laing (1943) administered

Wonderlic Personnel Test Forms D, E, and F to a sample of 118 individuals applying for work within an industrial plant, and to 71 employees of that same plant who were seeking a transfer to a better job. Ninety-seven men and ninety-two women comprised the total sample of 189 applicants. This sample ranged in age from 18 to 57 years, with a mean age for the group of 25 years. In terms of education, the sample ranged from 4 to 17 years of schooling, with the average individual having completed high school.

Sixty-one applicants took Form D; 72 applicants took Form E; and 56 applicants took Form F. Each applicant was given 12 minutes in which to answer as many test questions as he possibly could. Then note was made of the number of questions answered by each applicant at the end of 12 minutes. At this point each applicant was given an additional 12 minutes to work on the remaining unanswered test questions.

Pearsonian coefficients of correlation were computed between the 12-minute and 24-minute scores for each Wonderlic test form, with the following results reported: Form D--.947; Form E--.943; Form F--.923; and All Forms--.944.

While noting that the average increase in score for the 12-minute and 24-minute administrations were statistically significant for all three forms, Wright & Laing (1943) stated that "when the increase is interpreted in terms of the norms issued by the author, it does not appear to be so significant (p. 318)."

Therefore Wright & Laing (1943) concluded that:..

> the use of a longer time interval in the Wonderlic Personnel Test would result in a small increase in the

> obtained raw score. However, there is practically no change in rank from the use of the test in this manner. . . . (therefore) it would not appear economical to lengthen the administration of a test beyond that point at which it is effective (p. 319).

PRACTICE EFFECT

Hay (1952) administered Wonderlic Personnel Test Forms D and F in a counterbalanced fashion (rotating order) to 200 young women applicants for clerical positions within a large industrial organization in 1945. Form D followed by Form F was administered to 100 applicants, and Form F followed by Form D was administered to 100 applicants to neutralize any difference in difficulty between the two test forms. The results of this study showed that "there is a significant practice effect when one form of the Personnel Test is given immediately following the other (p. 345)." This practice effect reported amounted to 1.95 raw score points.

COMPARABILITY AND EQUIVALENCY OF WONDERLIC TEST FORMS

Hay (1952) also reported that there was a difference in the difficulty of Wonderlic Personnel Test Forms D and F, and that Form F was 1.36 score points easier on the average than Form D. This difference between test forms D and F was reported as being statistically significant at the .01 level, and was more than three times the sigma of the differences of the means. (The practice effect of 1.95 raw score points reported by Hay took into account an adjustment of scores based on this difference of 1.36 score points between forms.)

Weaver & Boneau (1956) challenged the statement made by the authors of the Wonderlic Personnel Test that the five test forms then available--A, B, D, E, and F--were equivalent forms and were inter-

changeable. A routine check of nearly 300 cases in which Form A and Form B had been administered at random to supervisory candidates within a large industrial firm showed that Form A seemed to be distinctly easier than Form B, the mean difference being 3.3 score points, which corresponded to a difference of from 10 to 20 centile units in the average range depending upon the particular test norms employed.

This suggestive finding led to a preliminary study utilizing 30 students enrolled in a psychology class. Half of the students were given Form A followed by Form B, and half were given Form B followed by Form A. The results of this preliminary study using a small sample showed that Form A was significantly easier than Form B at the .02 level of significance.

With two suggestive findings indicating that Form A was easier than Form B, Weaver & Boneau (1956) began a systematic study of the difficulty level of all five Wonderlic forms then available. A, B, D, E, and F.

Utilizing a Latin Square 5 x 5 design to equalize for practice and fatigue effects, Weaver & Boneau (1956) administered the five Wonderlic test forms to 70 students comprising two classes in psychology. Statistical analysis of the resulting data (see Table 1) indicated that the Wonderlic test forms fell roughly into two groups: A and B comprising a group of greater difficulty and higher variability than D, E, and F. "This accords with the history of the development of the test. Forms D, E, and F are made up of items selected from the Otis Higher, while A and B were developed later and include

TABLE 1

MEANS AND SD'S OF SCORES ON THE FIVE FORMS FOR 70 S's

Form	A	B	D	E	F
Mean	29.79	27.84	31.37	31.31	32.91
SD	6.01	7.39	5.73	5.57	5.78

Note --Reprinted from an article by Weaver & Boneau published in the 1956 Journal of Applied Psychology.

types of items not found in the Otis (pp. 127-128)."

Analysis of variance showed that individual differences and form differences were highly significant Nine of the ten form differences were statistically significant: two at the .02 level; three at the .01 level; and four at the .001 level. Only Wonderlic test forms D and E failed to differ significantly from each other. These form differences were tested individually by computing a t ratio based on the error variance.

Interform reliability coefficients were also computed between the first and second forms administered to each subject. The results showed remarkable close agreement with the finding of Wonderlic & Hovland (1939) among forms D, E, and F. Forms A and B could not be compared because of the lack of data from the test author regarding these two forms.

As a result of their findings in terms of differences in difficulty level among the five Wonderlic Personnel test forms, Weaver & Boneau (1956) concluded that the Wonderlic test forms A, B, D, E, and F could not be regarded as equivalent and are by no means interchangeable.

Furthermore, Weaver and Boneau (1956) also pointed out quite correctly that these form differences were not merely statistically significant but also were of great practical importance.

> The Wonderlic is very commonly used as a screening device and critical scores are set for employment and other purposes. If the forms are used interchangeably it would appear that meeting or failing to meet a given minimum score would in many cases depend on the particular form an applicant chanced to be given. This chance factor is as high as 5 points, which is 10% of the total

> possible score range, and in centile units of the published norms may amount to 37% of the distribution. The average interform difference, 2.34, is 4.7% of the total possible score range and in centile units may be as much as 20% of the distribution. In view of these variations in difficulty, separate tables of norms should be constructed for each form, or conversion formulas should be developed to permit comparisons of scores on different forms (p. 128).

It should be pointed out that the study by Weaver & Boneau (1956) used a college population rather than an industrial population for the testing of the difficulty level of the Wonderlic test forms, and that the Wonderlic was designed for testing adults in industrial testing situations. This assumed defect in their study led to a follow-up investigation by Kazmier & Browne (1961) in an industrial testing situation.

Subjects for the Kazmier & Browne study were 590 male applicants desiring to enter a large manufacturing company's industrial apprenticeship program involving such trades as tool making, die making, and pipe fitting. The education of these applicants ranged from completion of the eighth grade to completion of college, with a mean of 11.77 school years. Ages of the applicants ranged from 17 to 38 years, with a mean of 21.81 years.

The subjects were tested in sixteen sessions, with about 37 subjects being tested at each session. Five Wonderlic Personnel Test forms were administered to these applicants: A, B, D, E, and F. Every fifth man took Form A, every fifth man took Form B, etc. In total, 590 subjects were tested in this fashion and comprised the sample, with 118 subjects being administered one of the five test forms.

Twenty-nine of the subjects were between the ages of 30 and 38 years, and a correction factor of three score points was added to the net score for these subjects as suggested by Wonderlic in the test manual. However, the mean score for each test form was computed both before and after the corrections for age were made.

> The over-all significance of the observed differences among both mean uncorrected and corrected scores was measured through calculations of F ratios. The Duncan Multiple Range Test . . . was then used to test all differences, taken two at a time, at the 5% level of significance (p. 129).

The mean years of education also was computed by test form, and the significance of differences among those taking the five forms was determined by calculating the F ratio.

Analysis of variance for differences among scores on the five Wonderlic Personnel Test forms indicated that the observed differences were significant at the .01 level of confidence whether or not the scores for those 30 years of age or older are corrected (see Table 2). In fact correcting the scores for those 30 years of age or older had the effect of increasing the obtained differences between the means of pairs of test forms.

Form B differed from all the other Wonderlic test forms at the .01 level of significance whether or not corrections were made for age differences. In addition, Forms D and F differed significantly from each other at the .05 level of significance when corrections for age were made.

> That these differences among test forms can not be ascribed to a chance variation among the Ss in regard to their educational level was demonstrated by testing the significance of observed differences in educational

TABLE 2

MEAN SCORES AND VARIANCES BY PERSONNEL TEST FORM

(N = 118)

Form	A	B	D	E	F
Mean	20.62	16.79	19.47	19.66	21.13
S^2	51.55	39.64	34.28	45.29	35.68
Corrected Mean[a]	20.77	16.87	19.55	19.84	21.38
Corrected S^{2a}	55.54	42.46	37.51	48.50	45.82

[a] Scores corrected for age.

Note.--Reprinted from an article by Kazmier & Browne published in the 1959 Journal of Applied Psychology

> level for those taking the different forms of the test (p. 131).

The F ratio for differences in educational level was not significant at the .05 level being tested.

Kazmier & Browne (1961) concluded that Form B was more difficult than any of the other Wonderlic Personnel Test forms and that it should not be regarded as directly equivalent to any of the other four forms of the test, and that Form D could not be regarded as directly equivalent to Form F. "These findings are particularly pertinent, since Wonderlic suggests that when two forms of the test are to be used, the best combinations are A and B or D and F (p. 132)."

Although the order of difficulty among Wonderlic test forms in the Kazmier & Browne (1961) study are somewhat consistent with the Weaver & Boneau (1956) study an inspection of Tables 1 and 2 reveals these differences: (a) order of difficulty of test forms in increasing order (Weaver & Boneau)--F, D, E, A, B; (b) order of difficulty of test forms in increasing order (Kazmier & Browne)--F, A, E, D, B. Thus in both studies, Form F was found to be the easiest Wonderlic test form and Form B was found to be the most difficult Wonderlic test form, with no significant difference between test Forms D and E. The discrepancy between the two studies lies with the position of Form A.

DIFFERENCES IN MOTIVATION

Jennings (1953) initiated a study using Wonderlic Personnel Test Forms A and B to test whether there might be significant differences in motivation among individuals taking tests for purely research purposes or for actual promotion purposes.

Forty supervisors who volunteered to participate in a testing program for purely research purposes were divided into two groups of 20 individuals each. Both groups were then administered Wonderlic Personnel Test Form A.

Three months later, each group was administered Wonderlic Personnel Test Form B. However, Group 1 (Control Group) was encouraged to cooperate in the test taking venture for purely research purposes, while Group 2 (Experimental Group) was encouraged to cooperate for the purpose of giving management additional information for determining whom among them to promote to higher supervisory levels.

The results of this study showed:

> Whereas the differences in means and sigmas were not significant between the first and second testing for the Control Group 1, the Experimental Group 2, believing their performance at the second testing would affect their opportunity for promotion, increased their mean score almost seven points (p. 168).

In addition, Jennings found that the subjects within the Control Group generally maintained relative rank-positions, while individuals within the Experimental Group often changed rank-positions. Thus, Jennings concluded that the factor of motivation should definitely be controlled for when administering industrial aptitude tests for research purposes.

CHAPTER III

HYPOTHESES

For the most part the only psychological research studies that have been reported on some investigation of an internal characteristic of the Wonderlic Personnel Test have been those on equivalency of forms--namely, the studies by Hay (1952), Weaver & Boneau (1956), and Kazmier & Browne (1961). And all of these studies have been on the earlier developed forms A, B, D, E, and F. As was mentioned in Chapter I, Forms D, E, and F are no longer available to personnel executives. All of the other studies referred to in Chapter II have something to do with the external effect of certain characteristics of the Wonderlic Personnel Test. For example, the propriety of the 12-minute time limit (Wright & Laing, 1943) of the Wonderlic Personnel Test; the effect of sequence or practice effect in test taking using the Wonderlic test forms (Hay, 1952); and the factor of motivation when using the Wonderlic (Jennings, 1953).

This research has served its purpose however by pointing out the possibility that there might be other "defects" in the claims for the Wonderlic Personnel Test as purported to be valid by the test author in Wonderlic & Hovland (1939) and in the Wonderlic Personnel Test Manual (1966)--all of which were described in full detail in Chapter I. What is needed at this time is an application

of the research findings described in Chapter II to the most current forms--I, II, IV, V, and EM. In this section an attempt will be made to reiterate the research findings reported in Chapter II and to tie them in with the claims made for the Wonderlic by the test author in Chapter I, and finally to logically arrive at a hypothesis which will further give credence to the claims made by Wonderlic for his Personnel Test forms or will make them less believable.

Wright & Laing (1943) noted that there was a significant increase in the scores on all three Wonderlic forms used in that study when the standard time limit (12 minutes) was doubled. The question might be raised, What would have happened had there been an extension of time greater than 12 minutes?

More recently Dubin, Osburn, & Winick (1969) investigated the hypothesis that Negroes would benefit more than whites when opportunities were available for extra testing time. In this study subjects were paired for race, sex, grade level, and on a socioeconomic status index (SES) developed by the authors according to certain criteria. Subjects were then divided into two groups: those taking speeded tests and those taking power tests. A subject within one group was paired with a subject within the other group so that the hypothesis could be tested by comparing the speed-power test differences between the matched Negroes with the differences between the matched whites.

The results of this study showed that Negro and white improvement scores do not significantly differ. In addition, both socioeconomic status and the Race X SES interaction were unrelated to improvements resulting from extra testing time. Consequently, the authors

of this study concluded that administering tests without time limits does not favor Negroes.

A robust test of the finding of Dubin, Osburn, & Winick (1969) would be a situation where Negroes and whites were given the Wonderlic Personnel Test under both a timed condition and under an unlimited time condition. Therefore Hypothesis I is the following: Negroes will not benefit more than whites under an unlimited time administration (power) procedure for Wonderlic Personnel Test Form I. In addition, this hypothesis more readily lends itself to the question originated by Wright & Laing (1943): what happens if you liberally extend the time limit on the Wonderlic Personnel Test?

In Personnel Testing, Guion has stated that he is of the opinion that most of the total score variance on the Wonderlic is due to verbal comprehension. In fact, he states that "In general, beyond the verbal cognitive factors, items tapping other intellectual abilities seem rather scarce (p. 221)." Because there is current concern over the potential race bias in tests used for industrial personnel selection, and a heavy stress on the "culturally deprived" theory which states that Negroes have had less exposure to books, television, radios, etc., the question arises: Could the Negroes truly be "culturally deprived" in terms of language abilities and other verbal cognitive factors? Because the Wonderlic Personnel Test is believed to be heavily biased in terms of verbal comprehension questions, Hypothesis II is the following: Negroes will perform less well than whites on the verbal items contained within Wonderlic Personnel Test Form I under standard administration procedures

(12-minute time limit). If whites do in fact have an advantage over Negroes in the language area, it should show up under timed conditions.

Past studies (Hay, 1952; Weaver & Boneau, 1956; Kazmier & Browne, 1961) have shown beyond a questionable doubt that mean scores from the original Wonderlic Personnel Test forms (D, E, and F) differed significantly from those devised later (A and B). These cited studies also showed that there was some discrepancy among the original form D, E, or F: namely, that Form F was much easier than either Forms D and E. The conclusion of these cited studies was that these test forms, with the exception of forms D and E, certainly could not be considered as equivalent forms and used interchangeably. Weaver & Boneau (1956) went so far as to state that "separate tables of norms should be constructed for each form, or conversion formulas should be developed to permit comparison of scores on different forms (p. 128)." In 1961 Wonderlic did provide a Conversion Table in his test manual which would allow a comparison of scores on nine Wonderlic test forms. "By adding or subtracting the conversion score to the results of the first test comparative results can be obtained (p. 13)." Guion (1965) states however, that "The conversion factors were determined by averaging the deviations at the mean and at the 3 quartile points. This does not, of course, make the forms technically equivalent. . . . (p. 222)."

In the Conversion Table (Manual, 1966) it may be noted that no points need be added or subtracted to convert scores among three forms exclusively: Forms I, II, IV. Furthermore, Wonderlic states in the

test manual that for alternate forms "The best combinations are:. . . I and II, I and IV, and II and IV (p.2)." Therefore, Hypothesis III is the following: An examinee taking either Wonderlic Personnel Test Form I, II, or IV and then taking either of the other two alternate forms should make the same final score on both the first form taken and on any alternate forms taken.

This hypothesis will accurately estimate the efficacy of the Conversion Table and examine the three test forms which are considered to be exactly equivalent by Wonderlic. There should be a perfect positive (+1.00) correlation coefficient between test forms I and II, I and IV, and II and IV if they are truly equivalent. The guideline for comparability of test forms is specifically set forth in the Standards for Educational and Psychological Tests and Manuals:

> If two forms of a test are published, both forms being intended for possible use with the same subjects, the means and variances of the two forms should be reported in the test manual along with the coefficient of correlation between the two sets of scores. If necessary evidence is not provided, the test manual should warn the reader against assuming comparability.

If these alternate test forms are equivalent, we would expect to find a high positive correlation between them, along with similar means and standard deviations.

Wonderlic also states (Manual, 1966): "The forms A, B, I, II, IV, V, and EM have been so devised as to be equal in per cent of difficulty for units of ten questions, as well as for total score (p. 4)." Therefore, Hypothesis IV is the following: For Wonderlic Personnel Test Forms I, II, and IV, there is not an equality in

terms of per cent of difficulty for units of ten questions.

A summary of the four hypotheses advanced in this research project is given below:

1. Negroes will not benefit more than whites under an unlimited time administration (power) procedure for Wonderlic Personnel Test Form I.

2. Negroes will perform less well than whites on the verbal items contained within Wonderlic Personnel Test Form I under standard administration procedures (12-minute time limit).

3. An examinee taking either Wonderlic Personnel Test Form I, II, or IV and then taking either of the other two alternate forms should make the same final score on both the first form taken and on any alternate forms taken.

4. For Wonderlic Personnel Test Forms I, II, and IV, there is not an equality in terms of per cent of difficulty for units of ten questions.

CHAPTER IV

PROCEDURES AND RESEARCH DESIGNS

In this section a description will be given of the subjects, research designs, and statistical analysis procedures used in the testing of the four hypotheses. It should be pointed out that the subjects, designs, and statistical procedures for the testing of Hypotheses I and II are identical and are distinctly separate from the subjects, designs, and statistical procedures used in the testing of Hypotheses III, and IV.

SUBJECTS

All subjects in the research study were currently enrolled in an accredited Texas state-supported college during the summer of 1969. In addition all subjects included in the sample for each educational institution were required to meet the following criteria: (a) age--under 29 years of age, (b) sex--male or female, and (c) educational level--total completed college hours within the range of 25-90 semester hours.

These criteria are in accordance with established guidelines set forth in the Wonderlic Personnel Test Manual (1966), which states the following: (a) no correction factor points need be added to the raw score of individuals under 29 years of age; (b) the mean score for males with 2-3 years of college is 26.9, and the mean score

for females with 2-3 years of college is 25.7; and (c) the mean score for the educational group classified 1 year of college is 24.4, and the mean score for the educational group classified 2-3 years of college is 25.6.

The subjects for the research design for Hypothesis I and for Hypothesis II were drawn from two Texas educational institutions; (a) White subjects were drawn from Sam Houston State University, Huntsville, Texas, and (b) Negro subjects were drawn from Texas Southern University, Houston, Texas.

These two educational institutions were chosen specifically for the testing of Hypotheses I and II because of their comparability in terms of admission policies. Admission into Texas Southern University is based upon completion of a high school degree with no entrance test required. Admission into Sam Houston State University is based upon completion of a high school degree with no entrance test required, except for those students graduating in the lowest quartile of their high school graduation class. A cut-off score on the A.C.T. eliminates those high school students who are in the bottom 10% of the lowest quartile of their graduation class from admission to Sam Houston State University. Therefore, for all practical purposes the two colleges--Sam Houston State University and Texas Southern University--were deemed comparable in terms of admission policies.

Subjects at Sam Houston State University were enrolled in a summer six-week course in English 266; subjects at Texas Southern University were enrolled in a summer six-week course in English 231. Both English 266 and English 231 are considered to be comparable courses by

the registrars of these two colleges.

Two class sections of English 266 taught by a female instructor comprised the sample for the Sam Houston State University group. Two class sections of English 231 taught by a female instructor comprised the sample for the Texas Southern University group. Class sections at both schools were scheduled one after the other, and each section was a morning class lasting one hour and thirty minutes. Testing was completed within a three hour period at each of these schools, but on different days. Subjects at Sam Houston State University were tested by a White experimenter, while subjects at Texas Southern University were tested by a Negro experimenter.

For the testing of Hypotheses III, and IV, subjects were drawn from the following courses at the University of Houston, Houston, Texas: (a) two class sections of sophomore English 231, both taught by the same female instructor; (b) two class sections of sophomore English 238, both taught by a male instructor; and (c) two class sections of Psychology 133 (Introductory Psychology), both taught by the same male instructor.

These class sections lasted for one hour and thirty minutes, and all were daytime sections except for one early night section. Testing within these sections was completed within a four day period by a White experimenter.

Entire class sections were tested at each educational institution. Any individual tested who failed to meet the subject criteria was

excluded from the sample representative of that college. In addition no Negroes, Latin Americans, or foreign students were included in the sample representative of Sam Houston State University and the University of Houston. Thus, the Sam Houston State University and the University of Houston groups were exclusively composed of White individuals; and the Texas Southern University group was exclusively composed of Negro individuals.

TEST ADMINISTRATION PROCEDURES

Prior to testing, the instructor introduced the experimenter by name, college, and departmental affiliation. The instructor then informed the group that they were being allowed to participate in a "research project" and that they would be expected to give the experimenter their fullest cooperation.

The experimenter greeted the class and once again emphasized the fact that they would be participating in a "research project" involving some tests, and that their scores would be treated as confidential. The stressing of the word "research project" was deemed to satisfy the requirement of equal motivation for all groups, taking into account the finding by Jennings (1953).

The experimenter requested that the students form single rows, one student sitting directly behind another, and leave an unfilled row of chairs between each row of students. The students were informed that the only materials they would need would be several pieces of blank notebook paper, and that all other materials would be furnished by the experimenter.

Regular #2.5 lead pencils were then passed out by the experimenter

to the students, followed by a colored questionnaire form, and an Optical Scanner answer sheet. The colored questionaire forms were coded to represent the educational institution according to this format: (a) pink--Sam Houston State University, (b) green--Texas Southern University, and (c) yellow--University of Houston. (see List 1).

The following instructions were then given for filling out the colored questionnaire form and the Optical Scanner answer sheet respectively:

> 1. Fill in your Name, last name first, first name and middle initial. Where it says Course fill in _____ and Section number ____. Then write your age in a numerical fashion (ex: 18, 19, or whatever) in the blank labeled Age. Where it says Sex, fill in M for male or F for female. Where it says Total Completed College Hours, fill in the total number of college hours which you have successfully completed at present. Where it says Current GPA, fill in your current grade point average and indicate the point system that your school is on (ex: 2.3/4.0). Where it says College Enrolled in, fill in the college under which you are enrolled (ex: Arts and Sciences, Business, Engineering, etc.); and then fill in your Major, (ex: English, Mathematics, Finance, etc.). Then fill in the total number of hours that you have completed in English and in Mathematics, (ex: 0,3,6, etc.). Now pass the questionnaires to the person on the front of your row.
>
> 2. Now fill in your name--last name, first name and middle initial--in the space on the answer sheet labeled SCHOOL. Record your course number in the section labeled INSTRUCTOR. Record your class section number in the space labeled GRADE. Record the name of your school, using initials, in the space labeled CITY. (For Sam Houston State University and Texas Southern University only: Record Form 1 in the space labeled TEST).
>
> Now at the bottom right hand corner of this answer sheet, fill in the following information: in the space labeled STUDENT NUMBER, write down your student number and then fill in the corresponding blanks immediately below those numbers. Where it says SEX, fill in the blank B if you are a male and G if you are a female. Under the section labeled BIRTH DATE, fill in your

LIST 1

STUDENT QUESTIONNAIRE

Name:_____Course: English_____Section:_____
Last First M.

Age: _____ Sex:_____ Total Completed College Hours:____ Current GPA:____

College Enrolled in:_____ Major:_____

Completed Hrs. in English:_____ Completed Hrs. in Mathematics:_____

> numerical age at present, and fill in the appropriate blanks below these numbers. Do not put anything in the column labeled MONTH. For the space labeled GRADE, fill in blank three if you are a freshman, blank four if you are a sophomore, blank five if you are a junior, and blank six if you are a senior. If you do not fall into one of these classifications (ex: P.B. or graduate student), fill in blank seven. Now check to see that you have filled in all of the appropriate blanks; then pass the answer sheets up to the first person on the row.

After the answer sheets were collected, the experimenter then passed out the test forms: for the Sam Houston State University and Texas Southern University groups the test booklets were not distributed individually, as they were for the University of Houston group. The students were then instructed to place their name on the test booklet(s) and the date, and to wait for instructions before beginning the test.

Instructions for taking the test(s) were delivered to the group by the experimenter according to the Directions for Administration under a 12-minute time limit as set forth in the Wonderlic Personnel Test Manual (1966):

> "Read the first page carefully and work the sample problems. Do not turn the page until you are told to do so." After those being tested have had an opportunity to work each of the samples and read all the directions on the first page, the examiner should ask if everyone understands the way in which they are to work. He then says: "You will have exactly 12 minutes in which to answer as many problems as you can. You probably will not be able to finish all of them. Work as carefully as you can, but do not spend too much time on any one problem. I will stop you at the end of 12 minutes. Begin now!"(p. 2).

In addition to these instructions, the groups at the University of Houston were told that they had a packet of three test booklets stapled together, and that they were not to remove these staples

but were to work on the booklets in order. The students were also informed that the directions for taking the test would be given only once because the same instructions applied to all three tests. The students were instructed to place their name on the front page of each test booklet.

Each subject in the University of Houston group was given 12-minutes to complete each test. At the end of the first and second 12-minute periods the subjects within this group were told to open the next test booklet and begin working the problems. Upon completion of the third test booklet, the last person on each row was instructed to pass his packet of tests forward and each person was told to place their packet on top of the packet of tests taken by the person sitting directly behind them.

A somewhat different procedure was employed for the groups tested at Sam Houston State University and at Texas Southern University. At the end of the 12-minute testing period, each subject was asked to close his test booklet and to pass his pencil to the person sitting at the front of the row. A count of the pencils for each row was made by the experimenter as he then collected the black lead pencils. Then the experimenter distributed red lead pencils to the group, and issued the following instructions:

> You are now to reopen your test booklet and work on any of the problems that remain unanswered. You will have adequate time to complete the test. Time in whole minutes will be recorded on the front blackboard and note the time when you finish. Record this time in the upper right-hand corner on the front page of your test booklet. Begin now!

After all subjects in the Sam Houston State University and the Texas Southern University groups had completed the test, the students

were asked to record a "yes" or a "no" at the bottom of the front page of their test booklet as an answer to each of the following questions:

(a) Did you feel like you needed more time to finish the test?

(b) Did you feel motivated to do as well as you could on the test?

Then the experimeter collected the test booklets.

At the end of the testing period for each group a brief, general discussion was given by the experimenter concerning the objectives of the research study and then the class was dismissed.

The data recorded on the colored questionnaire forms provided the basis for determining which subjects tested could be included in the sample for each college. After selecting the subjects for each sample, the test booklet(s) for each subject was hand-scored using the Wonderlic Personnel test form scoring keys. Each test question was scored either right or wrong: omitted test items were scored as being wrong. Then the information from the colored questionnaire forms was transferred to the Optical Scanner answer sheet according to a specially-designed coding format. (see Table 3). Answers to the test items were then also transferred onto the Optical Scanner answer sheet for each subject according to this format: (a) A--right answer, and (b) B--wrong answer.

Answer sheet lines 98 and 99, respectively, were used to code the answer "yes" or "no" to the two questions posed to the Sam Houston State University and Texas Southern University groups. Line 100 was used to code the student's major for the Sam Houston State University and Texas Southern University groups according

TABLE 3

OPTICAL SCANNER FORM CODES

LINE	DATA RECORDED	CODE
1	Total Completed Hours in English	0-- 0 hours 2-- 8 hours 3-- 3 hours 5--11 hours 6-- 6 hours 7--12 hours 9-- 9 hours
2	Total Completed Hours in Mathematics	0-- 0 hours 3-- 3 hours 5-- 8-11 hours 6-- 6 hours 7--12-15 hours 8--16-18 hours 9-- 9 hours
3	Grade Point Average	0-- 1.5 or below 1-- 1.6 to 2.0 2-- 2.1 to 2.5 3-- 2.6 to 3.0 4-- 3.1 to 3.5 5-- 3.6 to 4.0
4	Race	0-- White 1-- Negro
5	Group	1-- Sam Houston State Univ. 2-- Texas Southern Univ 3-- Univ. of Houston
6	Latin Square Pattern (University of Houston only)	1-- I-II-IV 2--II-IV-I 3--IV-I-II
GRADE	Grade	3--under 30 semester hours 4- 30-60 semester hours 5--61-90 semester hours 6--91-120 semester hours

to this format: (a) A--Arts & Sciences, (b) B--Business, (c) E--Other; for Sam Houston State University: (a) C--Education, (b) D--Journalism; for Texas Southern University: (a) C--Pharmacy, (b) D--Industry.

Line 160 was used to code the student's major for the University of Houston group according to this format: (a) A--Arts & Sciences, (b) B--Business, (c) C--Engineering, (d) D--Education, (e) E--Other.

In this fasnion, information regarding the personal characteristics and the test-taking performance of each subject was transferred to each individual's answer sheet. Then the answer sheets were fed into the Optical Scanner machine and a tape was cut for submission to the computer for analysis.

RESEARCH DESIGN AND STATISTICAL ANALYSIS PROCEDURES FOR EACH HYPOTHESIS

Hypothesis I. Testing under the 12-minute (standard) time limit administration using black-lead pencils was termed the INITIAL STATUS phase. Testing under the unlimited (power) time limit administration using red-lead pencils was termed the FINAL STATUS phase.

Two scores--an INITIAL STATUS score and a FINAL STATUS score--were computed in the usual fashion for each individual included in the sample for the White (Sam Houston State University) and Negro (Texas Southern University) group. A GAIN score was computed for each subject in each of the two sample groups by taking the difference between the score obtained under the INITIAL STATUS phase and the score obtained under the FINAL STATUS phase. Thus,

for each subject in each of the two sample groups there were three scores: INITIAL STATUS, FINAL STATUS, GAIN. Also there was a score for the Time taken to finish the test for each subject under the FINAL STATUS phase.

The statistical procedures involved a computation of the mean and the standard deviation for the three test scores for each group of subjects. Then *t* tests for a difference between correlated means were made between the mean for the INITIAL STATUS and the mean for the FINAL STATUS for the White group and for the Negro group. And *F* ratios were computed to compare the variances under each of these conditions (Edwards, 1960, pp. 136-140).

Then *t* tests for a difference between uncorrelated means were made between the mean for the White group under the INITIAL STATUS phase and the mean for the Negro group under the INITIAL STATUS phase, and between the mean for the White group under the FINAL STATUS phase and the mean for the Negro group under the FINAL STATUS phase (Guildford, 1965, p. 183). And *F* ratios were computed to compare the variances under each of these two conditions.

Then a *t* test was computed to compare the difference between the uncorrelated means for the White and Negro groups under the GAIN condition in order to test the hypothesis. An *F* ratio was also computed to compare the variances of the two groups under the GAIN condition.

Hypothesis II. The design and statistical analysis procedure for Hypothesis II was identical to that for Hypothesis I with two exceptions: (a) the score for each subject was a partial score

computed only on the verbal items contained within the test as identified by the experimenter, and (b) in order to test the hypothesis, a t test was computed to compare the difference between the uncorrelated means for the White and Negro groups under the INITIAL STATUS phase. Additionally computed t tests only provided supplementary information about the groups.

The verbal items within Wonderlic Personnel Test Form I used in the computation of the verbal score for each subject in each group were identified as the following: Questions #2,3,4,5,6,7, 9,11,14,16,19,20,21,22,24,25,28,30,34,35,36,38,41,43,47,48. Thus a possible total score of 26 could be obtained by a subject. It might be pointed out that 17 of these questions were printed on the first page of the test booklet and should have been answered by almost every subject in view of the means obtained as reported in the section, Results. In total, these verbal items constituted 52% of the total number of items on the test.

These verbal items were independently identified by the experimenter and the Thesis Chairman. Both of these individuals arrived at the same conclusion in terms of which items should be labelled strictly verbal items.

Hypothesis III. A Latin Square paradigm (see Table 4) similar to the one employed by Weaver & Boneau (1956) was employed in testing this hypothesis. The testing paradigm uses a "unit" design for three subjects, and through this test distribution technique three groups were formed: Group A who took test booklets I-II-IV in that order; Group B who took test booklets II-IV-I in that order;

TABLE 4

PARADIGM FOR EQUALIZING PRACTICE AND FATIGUE EFFECTS AMONG WONDERLIC PERSONNEL FORMS I, II, IV

Order of Administration	Subject		
	1	2	3
First	I	II	IV
Second	II	IV	I
Third	IV	I	II

and Group C who took test booklets IV-I-II in that order. To insure that this pattern of test distribution remained intact, the experimenter individually handed each student a test packet. The first person on Row 1 received Forms I-II-IV stapled together; the second person on Row I received Forms II-IV-I stapled together; and the third person on Row I received Forms IV-I-II stapled together. Then the design was replicated throughout the rows in the classroom. During each testing session the first person handed a test packet received that test packet which succeeded according to the unit design the test packet handed to the last person tested in the immediately prior testing session.

Statistical analysis involved the computation of the mean and standard deviation for each test form, and the calculation of interform reliability coefficients. In addition a mean, standard deviation, and a confidence interval was computed for each test form within each group to examine the effect of order of administration and as a check on the practice effect noted by Hay (1952). Then t tests for differences between correlated means were computed between forms within each group. Correlation coefficients were also computed between test forms within a group, and a mean correlation coefficient was computed between paired test forms as a check on the overall interform correlation coefficient already computed. Calculation of the mean correlation coefficient was accomplished through transformation to z scores for averaging.

Hypothesis IV. Statistical analysis procedures for the testing of Hypothesis IV involved computing an average difficulty level

based on the number of subjects correctly answering the question divided by the number of subjects attempting to answer the question for blocks of ten questions by test form. Then *z* scores were computed to test for a difference between uncorrelated proportions for units of ten questions between test forms--*i.e.*, between the first ten questions on forms I, II, and IV, and *etc*.

CHAPTER V

RESULTS

In Chapter IV a discussion was presented that outlined the specific hypotheses to be tested in this research effort, and the design and statistical procedure appropriate for the testing of each hypothesis advanced. In this chapter a report of the statistical findings for each hypothesis has been presented in a likewise sequential manner.

For the testing of Hypothesis I and Hypothesis II two racial groups were administered Wonderlic Personnel Test Form I: a White group attending school at Sam Houston University and a Negro group attending school at Texas Southern University. These two groups were deemed by the experimenter to be comparable in age, sex, and total completed college hours (see List 2). The White group generally had completed one course in English more than had the Negro group. The Negro group generally had completed one or two more courses in mathematics than had the White group. The White group had an overall higher grade point average than did the Negro group, yet this may be an inflated picture of the true grade point average difference between the two groups because the experimenter obtained the grade point average on more than half of the Negro group from the Texas Southern University registrar's office. This

LIST 2

COMPOSITION OF THE SAMPLE ACCORDING TO PERSONAL AND ACADEMIC CHARACTERISTICS

WHITE:

(N = 35)

Age	Sex
18-- 3	M--14
19--11	F--21
20--12	
21-- 3	
22-- 2	
23-- 4	

Total College Hours

0-29-- 0	
30-60--26	
61-90-- 9	

Total College Hours in English

0-- 0
3-- 0
6-- 1
9--33
12-- 1

Total College Hours in Mathematics

0--11
3-- 6
6--15
9-- 3
12-- 0

NEGRO:

(N = 27)

Age	Sex
18-- 2	M--13
19--12	F--14
20-- 6	
21-- 2	
22-- 4	
23-- 0	
24-- 1	

Total College Hours

0-29-- 0
30-60--22
61-90-- 5

Total College Hours in English

0-- 0
3-- 0
6--24
9-- 3
12-- 0

Total College Hours in Mathematics

0-- 1
3-- 3
6--16
9-- 4
12-- 3

LIST 2

COMPOSITION OF THE SAMPLE ACCORDING TO PERSONAL AND ACADEMIC CHARACTERISTICS

WHITE:
(N = 35)

Major College Curriculum

Arts & Sciences--12
Business-- 5
Education--12
Journalism-- 4
Unspecified-- 2

Current Grade Point Average

0.0-1.5-- 0
1.6-2.0-- 6
2.1-2.5-- 6
2.6-3.0--15
3.1-3.5-- 5
3.6-4.0-- 3

Question #98

"Yes"-- 8
"No"--27

Question #99

"Yes"--16
"No"--19

NEGRO:
(N = 27)

Major College Curriculum

Arts & Sciences--13
Business-- 4
Education-- 3
Industries-- 3
Pharmacy-- 4

Current Grade Point Average

0.0-1.5-- 7
1.6-2.0--10
2.1-2.5-- 6
2.6-3.0-- 5
3.1-3.5-- 0
3.6-4.0-- 0

Question #98

"Yes"-- 2
"No"--25

Question #99

"Yes"--14
"No"--13

procedure was necessary due to the large number of omissions of grade point averages on the questionnaire forms by the Negro group. This procedure was not necessary for the Sam Houston University group because there were no omissions of grade point averages on the questionnaire forms by this group. Hence, the picture of the representation of grade point averages for the Negro group is more likely a realistic picture than is the one for the White group.

In regards to question #98--Did you feel like you needed more time to finish the test?--approximately 23% of the White group responded "yes", while only 1% of the Negro group responded "yes". Likewise there is a difference in the response pattern of the two groups in regards to question #99--Did you feel motivated to do as well as you could on the test?--in that approximately 54% of the White group responded "no", while only 49% of the Negro group responded "no". As was mentioned earlier in Chapter V great care was taken in the testing of both of these groups to insure that every subject had enough time to complete all of the test questions. The mean time under the FINAL STATUS condition for the two groups is as follows: White--15.00 minutes; Negro--16.52 minutes. Therefore, it appears that some of the subjects within the White group exhibited a great "need to achieve" as reflected in their response of "yes" to question #98, while at the same time denying that they were motivated to take the test as reflected in their "no" response to question #99.

The statistical results obtained in the testing of Hypothesis I supported that Hypothesis (see Table 5). There was a significant

TABLE 5

SIGNIFICANCE OF GAINS ON TOTAL TEST SCORES
ON WONDERLIC PERSONNEL TEST FORM I

Condition	Group						D	F	t
	White (N = 35)			Negro (N = 27)					
	Mean	V	SD	Mean	V	SD			
Initial Status	24.29	10.24	3.2	18.48	15.21	3.9	5.81	1.48	6.33**
Final Status	30.54	16.00	4.0	25.41	26.01	5.1	5.13	1.62	4.37**
Gain	6.25	4.84	2.2	6.93	10.24	3.2	.68	2.12*	.93[a]

[a] Corrected to account for significant variance differences

* $p < .05$

** $p < .01$

(at the .01 level) gain for the White and for the Negro group when given extra testing time. However, the overall difference between the White group and the Negro group under the GAIN condition was not significant, and thereby supported the hypothesis that Negroes will not benefit more than whites under an unlimited time administration procedure. This finding supports the results obtained in the study by Dubin, Osburn, & Winick (1969).

There is however, a significant (at the .01 level) difference between the White and the Negro group under both the INITIAL STATUS condition and the FINAL STATUS condition, with the Negro group obtaining a lower mean score and exhibiting greater variability than the White group under both conditions. In fact the Negro group under the FINAL STATUS condition obtained a mean score slightly higher than the mean score for the White group under the INITIAL STATUS condition. These results indicate that Wonderlic Personnel Form I tends to distinguish between the two racial groups.

Similar results were obtained in the testing of Hypothesis II in regards to the two groups. Both groups exhibited a significant (at the .01 level) gain between the INITIAL STATUS and the FINAL STATUS conditions. Likewise there was a significant (at the .01 level) difference between the two groups under the INITIAL STATUS and the FINAL STATUS conditions, with the significant difference under the INITIAL STATUS condition supporting the hypothesis that Negroes will perform less well than whites on the verbal items contained within Wonderlic Personnel Test Form I under standard administration procedures (see Table 6).

TABLE 6

SIGNIFICANCE OF GAINS ON VERBAL TEST SCORES
ON WONDERLIC PERSONNEL TEST FORM I

Condition	Group						D	F	t
	White (N = 35)			Negro (N = 27)					
	Mean	V	SD	Mean	V	SD			
Initial Status	14.86	3.61	1.9	11.59	5.29	2.3	3.27	1.46	6.03**
Final Status	17.29	4.84	2.2	15.22	6.25	2.5	2.07	1.29	5.39**
Gain	2.43	1.69	1.3	3.63	2.56	1.6	1.20	1.52	3.21**

** $p < .01$

There was also a significant (at the .01 level) difference between the White and the Negro group under the GAIN condition. This difference suggests that the Negro group was able to answer more of the verbal questions than the White group when given additional testing time. Thus it appears that the 12-minute time limit administration procedure for the Wonderlic Personnel Test penalizes the Negroes to some extent on the verbal items.

For testing Hypothesis III a mean and a standard deviation were computed for Wonderlic Personnel Test Form I, II, and IV based on the University of Houston sample (N=108). Kuder-Richardson 20 reliability estimates and the mean difficulty of test items by test form were also computed. Then correlation charts between alternate test forms were plotted and interform correlation coefficients were computed (see Table 7).

There were no significant differences between Wonderlic Personnel Test Forms I, II, and IV in terms of mean score and variability, or in the Kuder-Richardson 20 reliability estimates or in mean difficulty of test items. In terms of these statistics, Forms II and IV appear to be almost identical, while Forms I and IV appear to be most dissimilar even though this dissimilarity is slight.

In terms of the interform correlation coefficients these alternate test forms do not, however, appear to be similar, comparable and equal. Regardless of the parallelism between all of the statistics mentioned in the above paragraph, an examination of the correlation charts presented in Appendix A shows the high degree of variability exhibited by subjects when taking alternate Wonderlic Personnel Test Forms I, II, and IV. Correlation Chart I-II shows that the most variable

TABLE 7

SUMMARY OF STATISTICAL DATA BY TEST FORM
(N = 108)

Statistical Data:	Wonderlic Personnel Test Form		
	I	II	IV
Mean	27.35	28.00	28 02
Standard Deviation	4.78	5.58	5.62
K-R 20 r_{tt}	.756	.795	.797
Mean Difficulty of Test Items	.547	.560	.560

Interform Correlation Coefficients	I-II	I-IV	II-IV
	.621	.573	.634

subject taking these two alternate forms exhibited a discrepancy of 15 points, or 30% of the total test variance, between these two forms. Correlation Chart I-IV shows that the most variable subject taking these two alternate forms exhibited a discrepancy of 14 points, or 28% of the total test variance, between these two test forms. And Correlation Chart II-IV shows that the most variable subject taking these two alternate test forms exhibited a discrepancy of 13 points, or 26% of the total test variance, between these two test forms. Clearly these most variable subjects are not making the same score on alternate test forms I, II, and IV. Nor are many other subjects, as an inspection of these correlation charts reveals. Only 14 subjects made identical test scores on forms I and II; only 15 subjects made identical test scores on forms II and IV; and only 27 subjects made identical test scores on forms I and IV.

In Standards for Educational and Psychological Tests and Measurements reliability is defined clearly and concisely: "Reliability refers to the accuracy (consistency and stability) of measurement by a test. . . . the terms 'consistency' and 'stability' are needed to describe, respectively, form-associated and time-associated reliability." Clearly these Wonderlic test forms are not statistically reliable in terms of total test score for the majority of the sample when stability is taken into account.

As a further check on the low interform correlation coefficients obtained, a mean and a standard deviation was computed for each test form within each group to examine the effect of order of administration. Then a confidence interval was computed for each form mean

within each group (See Table 8). The results obtained showed that there was an increase of 2.7 score points in Group A between the first and second test administered; an increase of 3.3 score points in Group B between the first and second test administered; and an increase of 2.3 score points in Group C between the first and second test administered. To test for the significance of these obtained differences in means between the first and second test administered within each group, t tests were computed for correlated means (See Table 9). The t tests showed that there was a significant (at the .01 level) difference between the mean for the first test administered and the second test administered within each group. This finding indicates that there is a considerable "warm-up" or 'practice" effect operating within each group between the first and second tests administered, and is greater than the practice effect of 1.95 points noted by Hay (1952). Additional t tests between the second and third test administered showed that there was not a significant difference between the means of these two tests for Group A and Group C. There was a significant (at the .01 level) decrease in mean score between the second and third test administered within Group B. Therefore no concrete conclusion can be reached concerning the interaction effect of tests two and three within each group.

Because the Latin Square paradigm used did not completely partial out the practice effect, the low intercorrelation coefficients obtained between paired test forms become questionable. So correlation coefficients were computed between paired test forms within each group (See Tables 10, 11, and 12) and a mean correlation coefficient between paired test forms was computed (See Table 13).

TABLE 8

CONFIDENCE INTERVALS (.05) FOR TEST FORM MEANS
BY ORDER OF ADMINISTRATION

Group	Mean	S.D.	Confidence Interval
A: I-II-IV			
I	25.7	5.0	24.07-27.33
II	28.4	6.2	26.38-30.42
IV	29.3	6.4	27.20-31.40
B: II-IV-I			
II	26.1	5.0	24.31-27.89
IV	29.4	4.5	27.79-31.01
I	28.3	4.9	26 55-30.05
C: IV-I-II			
IV	25.8	4.8	24.29-27 31
I	28.1	4.0	26.95-29.25
II	29.1	4.9	27.56-30.64

TABLE 9

t-TESTS FOR DIFFERENCES BETWEEN MEANS BY ORDER OF ADMINISTRATION

Group	*t*	*df*	Probability
A: I-II-IV			
I-II	3.82	36	less than .01
I-IV	4.89	36	less than .01
II-IV	1.36	36	-
B: II-IV-I			
I-II	3.20	30	less than .01
I-IV	5.39	30	less than .01
II-IV	17.94	30	less than .01
C: IV-I-II			
I-II	1.60	39	-
I-IV	10.98	39	less than .01
II-IV	16.06	39	less than .01

TABLE 10

CORRELATION COEFFICIENTS FOR ORDER OF ADMINISTRATION
GROUP A: I-II-IV

(N= 37)

Form	I	II	IV
I	-	.73	.73
II	.73	-	.80
IV	.73	.80	-

TABLE 11

CORRELATION COEFFICIENTS FOR ORDER OF ADMINISTRATION
GROUP B: II-IV-I

(N= 31)

Form	II	IV	I
II	-	.78	.71
IV	.78	-	.72
I	.71	.72	-

TABLE 12

CORRELATION COEFFICIENTS FOR ORDER OF ADMINISTRATION
GROUP C: IV-I-II

(N= 40)

Form	IV	I	II
IV	-	.57	.65
I	.57	-	.63
II	.65	.63	-

TABLE 13

MEAN CORRELATION COEFFICIENTS BY TEST FORMS

Correlation Coefficient	Group A	Group B	Group C	Mean
r_{12}	.73	.71	.63	.69
r_{14}	.73	.72	.57	.68
r_{24}	.80	.78	.65	.75

The mean correlation coefficient obtained between Form I and Form II was .69, an increase from .62 as reported in Table 8. For Form I and Form IV the mean correlation was .68, an increase from .57. And for Form II and Form IV the mean correlation was .75, an increase from .63. Yet these correlations are still too low to insure adequate reliability in a test-retest situation using alternate test forms. And the correlations obtained even after partialling out the practice effect are still below the range of .82 to .94 as claimed by Wonderlic (Manual, 1966) for test-immediate retest situations.

In view of the fact that the Wonderlic test manual (1966) does not specifically state the correlation coefficient between Form I and Form II, between Form I and Form IV, and between Form II and Form IV, comparability of these Wonderlic test forms can not be assumed as outlined in Standards for Educational and Psychological Tests and Manuals. Due to the low interform correlation coefficients obtained even when accounting for the practice effect noted, Wonderlic Personnel Test Forms I, II, and IV can not be regarded as equal, equivalent, or alternate test forms. Hypothesis III is thereby supported.

In testing Hypothesis IV, the average difficulty level of blocks of ten questions was computed according to the ratio: $\frac{\text{number of Ss passing}}{\text{number of Ss attempting}}$, for each test form (See Table 14). Then z scores were computed to test for a difference between uncorrelated proportions for units of ten questions between forms, except for questions #41-50 where the N on each form was too small to permit computation.

TABLE 14

AVERAGE DIFFICULTY (PROPORTION CORRECT) OF BLOCKS OF TEN QUESTIONS BY TEST FORM

Question Group	Form I	Form II	Form IV
Questions # 1-10	.92 (N=108)	.91 (N=108)	.94 (N=108)
Questions #11-20	.82 (N=108)	.76 (N=108)	.80 (N=108)
Questions #21-30	.77 (N=100)	.75 (N=104)	.61 (N=105)
Questions #31-40	.56 (N=50)	.57 (N=67)	.59 (N=77)
Questions #41-50	.39 (N=7)	.36 (N=14)	.36 (N=17)

The statistical results (See Table 15) indicated a significant difference exists between all paired test forms within question block three (questions #21-30) and question block four (questions #31-40). All of the reported means for each test form by group as reported in Table 8 fall within question block three, so there is evidence that most subjects were able to attempt to solve the problems within question block three (as also noted by the N's reported in Table 14). Therefore, the hypothesis that for Wonderlic Personnel Test Forms I, II, and IV there is not an equality in terms of per cent of difficulty for units of ten questions is supported.

TABLE 15

z-SCORE RESULTS FOR SIGNIFICANCE OF DIFFERENCES BETWEEN PAIRED TEST FORM PROPORTIONS ON SUCCESSIVE GROUPS OF TEN QUESTIONS

Question Group	I-II	I-IV	II-IV
Questions # 1-10	.27	.61	.84
Questions #11-20	1.07	.54	.71
Questions #21-30	1.67*	14.10**	12.73**
Questions #31-40	5.55**	1.80*	1.82*
Questions #41-50	-	-	-

*p < .05
**p < .01

CHAPTER VI

SUMMARY AND CONCLUSIONS

The purpose of this study was to scrutinize a very popular and widely used mental ability or general intelligence test, the Wonderlic Personnel Test, in a very rigorous fashion in relation to three general questions: (a) Does the Wonderlic Personnel Test favor one ethnic group over another ethnic group to any great extent? (b) If the Wonderlic Personnel Test does in fact favor one ethnic group over another ethnic group, what might be the basis for this differentiation? (c) Are the claims made by the test author of the Wonderlic Personnel Test in regards to its internal characteristics valid?

In order to answer the first question, a sample of White college students (N=35) and a sample of Negro college students (N=27) were administered Wonderlic Personnel Test Form I under both the standard administration time limit of 12 minutes and under an unlimited (power) administration procedure. Each group was composed of summer college students of both sexes who were under 25 years of age and who had obtained 30-90 semester college hours. A comparison of the mean score for each group indicated that the Negro group scored considerably lower on Wonderlic Personnel Test Form I than the White group under both the standard time limit administration procedure and under the unlimited (power) administration

procedure. Both groups showed a significant gain in mean score under the power administration procedure. However, the gain in mean score obtained between the two groups under the additional time administration procedure was not significant and supported the hypothesis that Negroes will not benefit more than whites on Wonderlic Personnel Form I when given additional testing time. In fact, the mean score for the Negroes under the power administration procedure slightly surpassed the mean score for the Whites under the 12 minute administration procedure. Clearly, then, the answer to the first question is: Yes, the Wonderlic Personnel Test (Form I) does favor Whites over Negroes.

The second question was, "If the Wonderlic Personnel Test does favor one ethnic group over another, what might be the basis for this phenomenon?" To answer this question mean scores were computed for both the Whites and the Negroes based upon test items whose content was verbal in nature. The result of this procedure showed that the Negroes had a significantly lower mean score than did the Whites under both the 12-minute administration procedure and under the unlimited (power) administration procedure. Both groups showed a significant gain in mean score under power administration. However, the Negroes showed a significantly larger gain in mean score under the power administration than did the Whites. This finding answered the above question and supports the hypothesis that Negroes will perform less well than Whites on the verbal items contained within Wonderlic Personnel Test (Form I) under a standard administration time limit. It appears as though the verbal items within Wonderlic

Personnel Test Form I constitute a major type of item content that distinguishes between the two ethnic groups.

The finding that Negroes do not perform as well as Whites on Wonderlic Personnel Test Form I under both the standard time administration procedure and under an untimed administration procedure for both total test score and for a part-score based on verbal items clearly indicates one of two functions. Either there is a real difference between the two races in terms of mental ability level due to genetic factors, or there is only an apparent difference between the two races due to a difference in environmental, educational and cultural opportunities. It was not, however, the purpose of this study to resolve the question of why there is a difference between the two races in terms of an overall mental ability level, and at the time of the writing of this paper this question has not been conclusively resolved by any research efforts.

Yet the fact that there is a difference between Whites and Negroes in terms of their test taking performance on the Wonderlic Personnel Test clearly indicates that this test should not be administered to Negroes when the published norms are based exclusively on a White population. Furthermore, the usage of White norms for the Wonderlic Personnel Test when interpreting scores made by Negroes would be in violation of the Civil Rights Act of 1964, Title VII, as is pointed out in an article by Ash (1966) entitled "The Implications of the Civil Rights Act of 1964 for Psychological Assessment in Industry":

> But it seems clear that a test may be held to

> discriminate in the forbidden sense--i.e., to deny equal opportunity for consideration--under a number of other circumstances. The heart of the criticism of testing in this regard is that tests "screen out from opportunities for advancement those individuals from a background of cultural deprivation who because of this deprivation give an inferior performance on tests (Brim, 1965)," but not necessarily on the tasks for which the tests are supposed to be predictive.
>
> A test may operate to discriminate in the forbidden sense (a) when scores on it tend to differentiate between identifiable subgroups, where the subgrouping itself is not a relevant selection factor. . . .

In regards to a company using its own norms on the Wonderlic to interpret a score made on that test by a Negro, a word of caution supplied by the Department of Labor in the Federal Register (1968) to contractors and subcontractors can serve as a guideline for the company's policy:

> It is directed that each agency require each contractor regularly using tests to select from among cardidates for hire, transfer or promotion to jobs other than professional, technical, or managerial occupations . . . to have available for inspection, within a reasonable time, evidence that the tests are valid for their interded purposes. Such evidence shall be examined in compliance reviews for indications of possible discrimination, such as instances of higher rejection rates for minority candidates than nonminority candidates.
>
> Evidence of a test's validity should consist of empirical data demonstrating that the test is predictive of or significantly correlated with important elements of work behavior comprising or relevant to the job(s) for which candidates are being evaluated.
>
> Where there are data suggesting that such unfair discrimination exists (e.g., differential rates of rejecting applicants from different ethnic groups or disproportionate representation of some ethnic groups in employment in certain classes of jobs), then the contractor may be called upon to present evidence concerning the validity of his unscored procedures as well as of any tests which may be used. . . .

In regards to claims made by the test author, two aspects of the Wonderlic Personnel test were examined: (a) equivalency

of forms, and (b) an equality in terms of per cent of difficulty for units of ten questions between test forms. To examine these two internal characteristics of the Wonderlic Personnel Test, three alternate test forms judged by the test author to be equivalent were chosen for testing purposes: Forms I, II, and IV. Two reasons entered the decision to use these particular test forms: (a) these forms are claimed by the test author to be among the best forms to use for alternate testing purposes (test-retest), and (b) these forms are the only forms in the Conversion Table provided in the test manual for which there are no score corrections between test forms. Theoretically therefore, if any of the Wonderlic Personnel Test forms are truly equivalent test forms, then Forms I, II, and IV should best fit these qualifications.

In order to verify the internal characteristics claimed for the Wonderlic Personnel Test by its author, Wonderlic test forms I, II, and IV were administered to a sample of White summer school students (N-108) in a fashion--Latin Square paradigm--that would presumably take any practice or fatigue effects into account. Interform correlation coefficients were then computed between paired test forms, and the results showed a range of +.57 to +.63. Further treatment of the data examined for a practice effect or fatigue effect, and a practice effect was noted. Therefore, mean correlation coefficients between paired test forms were computed which accounted for the practice effect. The results produced estimates of interform correlation coefficients of between +.68 to +.75. These obtained mean interform correlation coefficients are below

the +.82 to +.94 range given in the Wonderlic Personnel Test Manual (1966) however. And because there was not a perfect correlation between Wonderlic Personnel Test Forms I, II, and IV and there was a high degree of variability among the students taking these test forms (See Appendix A), Hypothesis III was thereby supported.

The statistical treatment of the data to check on per cent of difficulty for units of ten questions between Wonderlic Personnel Test Forms I, II, and IV indicated that these test forms were not truly equivalent in the third and fourth blocks of ten questions. Therefore the hypothesis that Wonderlic Personnel Test Forms I, II, and IV are not equal in terms of per cent of difficulty for units of ten questions was supported.

In summary, the results of this study are at odds with the information found in the Wonderlic Personnel Test Manual (1966) and cast doubt on the adequacy of the interform reliability of these tests. Replication of this study in additional samples to test the generalizability of estimates of interform reliability of the Wonderlic should be performed.

BIBLIOGRAPHY

Ash, P. The implications of the Civil Rights Act of 1964 for psychological assessment in industry. American Psychologist, 1966, 21, 797-803.

Berdie, R. The ad hoc committee on the social impact of psychological assessment. American Psychologist, 1965, 20, 143-146.

Buros, O. (Ed.) The third mental measurements yearbook. New Brunswick, N. J.: Rutgers Univ. Press, 1949.

Buros, O. (Ed.) The sixth mental measurements yearbook. Highland Park, N. J.: Gryphon, 1965.

Dubin, J., Osburn, H., & Winick, D. Speed and practice: Effects on Negro and white test performances. Journal of Applied Psychology, 1969, 33, 19-23.

Edwards, A. Statistical analysis. New York: Holt, Rinehart and Winston, 1960.

Federal Register, 33, No. 186 (September 24, 1968). Validation of employment tests by contractors and subcontractors subject to the provisions of Executive Order 11246.

Guilford, J. P. Fundamental statistics in psychology and education. (4th ed.) New York: Mc-Graw-Hill, 1965.

Guion, R. M. Personnel testing. New York: McGraw-Hill, 1965.

Hay, E. Some research findings with the Wonderlic Personnel Test Journal of Applied Psychology, 1952, 36, 344-345.

Hovland, C. I., & Wonderlic, E. F. A critical analysis of the Otis Self-Administering Test of Mental Ability--Higher Form. Journal of Applied Psychology, 1939, 23, 367-387.

Jennings, E. The motivation factor in testing supervisors. Journal of Applied Psychology, 1953, 3, 168-169.

Kazmier, L., & Browne, C. G. Comparability of Wonderlic test forms in industrial testing. Journal of Applied Psychology, 1959, 43, 129-132.

Lindquist, E. F. Design and analysis of experiments in psychology and education. Boston. Houghton Mifflin, 1956.

McNemar, Q. Psychological statistics. (2nd ed.) New York: Wiley, 1955.

Standards for Educational and Psychological Tests and Manuals. Washington: The American Psychological Association, 1966.

Weaver, H. G., & Boneau, C. A. Equivalence of forms of the Wonderlic Personnel Test: A study of reliability and interchangeability. Journal of Applied Psychology, 1956, 40, 127-129.

Wonderlic, E. F. A selected, annotated bibliography for the Wonderlic Personnel Test. Northfield, Ill.: E. F. Wonderlic & Associates, Inc., 1966.

Wonderlic, E. F. 1967 Supplement to a selected, annotated bibliography for the Wonderlic Personnel Test. Northfield, Ill.: E. F. Wonderlic & Associates, Inc., 1967. (a)

Wonderlic, E. F. Wonderlic Personnel Test Manual. Northfield, Ill.: E. F. Wonderlic & Associates, Inc., 1967. (b)

Wonderlic, E. F., & Hovland, C. I. The Personnel Test: A restandardized abridgement of the Otis S-A test for business and industrial use. Journal of Applied Psychology, 1939, 23, 685-702.

Wright, J. H., & Laing, D. M. The time factor in the administration of the Wonderlic Personnel Test. Journal of Applied Psychology, 1943, 27, 316-319.

APPENDIX A

CORRELATION CHARTS

CORRELATION CHART I - II

$d_{x\ I}$

$d_{y\ II}$ \ x	14 15	16 17	18 19	20 21	22 23	24 25	26 27	28 29	30 31	32 33	34 35	36 37	38 39
45 44													
43 42													
41 40													
39 38								1				1	
37 36						1		1	2	3	1		1
35 34				1				2	1	4	1	1	
33 32					2	2	1		1	1	5		
31 30					1	2	4	1	1	3			
29 28						2	3	3	1		2		
27 26					2	2		3	2	1	1	1	
25 24				3	1	3	5	2	1				
23 22				2	3	6	1	3					
21 20				1	1		2						
19 18						1							
17 16	1				1								
15 14	1			1	1								

CORRELATION CHART 1 - IV

d_x I

d_y IV

y \ x	14 15	16 17	18 19	20 21	22 23	24 25	26 27	28 29	30 31	32 33	34 35	36 37	38 39
45 44													
43 42													1
41 40								1		1			
39 38									1				
37 36									1		4	1	
35 34						1	2	1	1	2	2		
33 32				1	1	2		1		6	1		
31 30				1		2	2	2	1	2			
29 28					3	1	4	3	2		1	2	
27 26						2	5	2	1		1		
25 24				3	1	5	1	2	1	1			
23 22	1			2	3	2	2	3	1		1		
21 20				1	3	2		1					
19 18						1							
17 16	1				1								
15 14						1							

CORRELATION CHART II - IV

d_x II

d_y IV

y \ x	14 15	16 17	18 19	20 21	22 23	24 25	26 27	28 29	30 31	32 33	34 35	36 37	38 39
45 44													
43 42												1	
41 40											1		1
39 38												1	
37 36							2	1		1	1	1	
35 34							1	1	1	2	3	1	
33 32							1	1	3	1	4	2	
31 30						1	2	3	1	1	1	1	
29 28				2	1	2	3	2	3	2			1
27 26				1	2	2		2	2	1		1	
25 24	1		1		5	3	1	1	1	1			
23 22		1		1	3	5			1	3		1	
21 20		1			3	1	2						
19 18						1							
17 16	2												
15 14					1								