The relation between measured intelligence and the ability to learn

Robert Stephen Peddicord
THE RELATION BETWEEN MEASURED INTELLIGENCE
AND THE ABILITY TO LEARN

by

Robert Stephen Peddicord

Approved:

[Signatures]

LIBRARY
UNIVERSITY OF RICHMOND
VIRGINIA
THE RELATION BETWEEN MEASURED INTELLIGENCE
AND THE ABILITY TO LEARN

by
Robert Stephen Peddicord

A thesis submitted in partial fulfillment
of the requirements for the degree
of Master of Arts in Psychology in
the Graduate School of the University of Richmond

July 1968
ACKNOWLEDGEMENTS

The writer is indebted to the Chairman of his Thesis Committee, Dr. W. H. Leftwich, for his helpful guidance. The writer is, likewise, indebted to the members of his Thesis Committee, Drs. H. C. Gwaltney and L. J. Tromater. Appreciation is also expressed to Mr. Peter Bahler, Director of the Computer Center, for his assistance in "debugging" computer programs written by the author.
## TABLE OF CONTENTS

<table>
<thead>
<tr>
<th>CHAPTER</th>
<th>PAGE</th>
</tr>
</thead>
<tbody>
<tr>
<td>I. Background and Statement of the Problem</td>
<td>1</td>
</tr>
<tr>
<td>II. The Method</td>
<td>30</td>
</tr>
<tr>
<td>III. The Results</td>
<td>47</td>
</tr>
<tr>
<td>IV. The Discussion</td>
<td>67</td>
</tr>
<tr>
<td>V. Summary and Conclusions</td>
<td>81</td>
</tr>
<tr>
<td>References</td>
<td>87</td>
</tr>
<tr>
<td>Appendix I: Item Statistics for the &quot;Listening Comprehension Tests&quot; Based Upon the Item Pretest Sample</td>
<td>92</td>
</tr>
<tr>
<td>Appendix II: Part A--Transcripts of &quot;Listening Comprehension Tests&quot;</td>
<td>96</td>
</tr>
<tr>
<td>Part B--&quot;Listening Comprehension Tests&quot; Answer Sheets</td>
<td>105</td>
</tr>
<tr>
<td>Appendix III: Part A--Description of I.B.M. 1620 Program Employed</td>
<td>110</td>
</tr>
<tr>
<td>Part B--Fortran Programs Written by the Author and Used in the Data Analysis</td>
<td>111</td>
</tr>
<tr>
<td>Vita</td>
<td>112</td>
</tr>
</tbody>
</table>
# LIST OF TABLES

<table>
<thead>
<tr>
<th>Table</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>I.</td>
<td>Means, Mean Item Test r's, and Internal-Consistency Reliabilities of the Three Alternate Forms of the &quot;Listening Comprehension Tests&quot; for the Item Pretest Sample</td>
<td>35</td>
</tr>
<tr>
<td>II.</td>
<td>Means and Standard Deviations of Ability Test Scores for the Sample in the Experiment</td>
<td>37</td>
</tr>
<tr>
<td>III.</td>
<td>Means and Standard Deviations of Differential Aptitude Test Scores for the Eighth Grade Students of the National Sample</td>
<td>39</td>
</tr>
<tr>
<td>IV.</td>
<td>The Means and Standard Deviations of the Weighted &quot;Listening Comprehension Tests&quot; Scores</td>
<td>48</td>
</tr>
<tr>
<td>V.</td>
<td>Single Classification Analysis of Variance (Repeated Measures) Between Log Transformed Weighted Scores of the &quot;Listening Comprehension Tests&quot;</td>
<td>50</td>
</tr>
<tr>
<td>VI.</td>
<td>Duncan Procedure for Testing Differences Between Log Transformed Weighted Total Scores on the &quot;Listening Comprehension Test&quot;</td>
<td>51</td>
</tr>
<tr>
<td>VII.</td>
<td>Guide to Abbreviations Used</td>
<td>52</td>
</tr>
<tr>
<td>VIII.</td>
<td>7 x 7 Intercorrelation Matrix for Raw Test Scores (N = 77)</td>
<td>55</td>
</tr>
<tr>
<td>IX.</td>
<td>Intercorrelations Between Measures of Gain (N = 76)</td>
<td>57</td>
</tr>
<tr>
<td>X.</td>
<td>9 x 9 Intercorrelation Matrix for Weighted Absolute Gain (N = 77)</td>
<td>59</td>
</tr>
<tr>
<td>XI.</td>
<td>9 x 9 Intercorrelation Matrix for Log Transformed Proportion of Possible Gain (N = 76)</td>
<td>61</td>
</tr>
<tr>
<td>XII.</td>
<td>9 x 9 Intercorrelation Matrix for Residual Gain (N = 77)</td>
<td>63</td>
</tr>
<tr>
<td>XIII.</td>
<td>Multiple Correlation Coefficients Between Residual Gain 1 and Residual Gain 2 Considered as Predictors, and Measures of Ability as the Criterion</td>
<td>66</td>
</tr>
<tr>
<td>FIGURE</td>
<td>PAGE</td>
<td></td>
</tr>
<tr>
<td>--------</td>
<td>------</td>
<td></td>
</tr>
<tr>
<td>1. Mean Weighted Scores on the &quot;Listening Comprehension Tests&quot;</td>
<td>49</td>
<td></td>
</tr>
</tbody>
</table>
CHAPTER I

BACKGROUND AND STATEMENT OF THE PROBLEM

There are perhaps as many definitions of intelligence as there are definers. However, one definition which has extensive commonsense appeal, according to Simrall (1947), identifies intelligence with the ability to learn, or to profit from experience. Certainly, many practicing psychologists have interpreted intelligence test scores as reflective of an individual's learning ability or potential for future learning. In his address as retiring President of the American Psychological Association, Garrett commented that: "It is undoubtedly true that intelligence involves the ability to learn . . ." (Garrett, 1946, p. 372).

Similarly, Simrall has commented:

With competence and ingenuity man solves the problems of everyday life. How except by previous learning, could he have achieved this ability? Surely no one is naive enough to suppose that a man is born with the skills necessary for the solution of any complex problem. What can be meant by the term 'intelligence' unless that meaning is ability to learn, to profit from previous experience? Is it plausible that intelligent behavior should be achieved except through experience? (Simrall, 1947, p. 27).
Learning is the modification of behavior through experience. More specifically, it is the improvement (with respect to a definite criterion) which occurs with practice (McGeogh, 1942, p. 3). Learning is thus measured in terms of gain due to practice (Simrall, 1947, p. 29). Therefore, if intelligence is to be defined as the ability to learn and if our intelligence tests do indeed measure "intelligence," then performance on a test of intelligence should be related to gains due to practice of tasks similar to those included in the intelligence test. The logic behind such reasoning is rather straightforward. Intelligence tests are standardized samples of behavior. Although behaviors included in the sample are generally not measures of learning, as such, but rather measures of achievement, it is assumed that the present level of achievement or performance on the test reflects past learning. Degree of proficiency of past learning is taken to be predictive of learning ability or probable level of proficiency in future learning. Thus, present achievement reflects previous gain in performance which is assumed to be predictive of future gains in performance.

Before proceeding with research relevant to the main question of the relation between measured intelligence and learning ability, it may be advisable to consider alternative explanations which may account for present levels of achievement. Present achievement adequately reflects past learning.
insofar as everyone has had an opportunity to learn under the same conditions. Consider, for example, an extreme case. If an English achievement test were administered to a French child, who had not been exposed to English, the test would probably not be indicative of his potential for learning English. In such a case an achievement test is an inappropriate measure of the ability to learn English. One may wonder, therefore, how appropriate our measures of intelligence which tap achievement are for predicting learning potential when subjects tested come from various environmental backgrounds having had differential learning experiences. Thus, intelligence, the attribute or process which intelligence tests supposedly measure, may be properly defined at least in part as learning ability, however, our achievement type tests of intelligence may not measure it for all or even most individuals. Only if the individual has had a maximum opportunity to learn, can present achievement be said to reflect adequately his learning potential.

Another way of viewing the problem of the relationship between present achievement and past learning is to consider present achievement as a measure of efficiency of storage of past learning. It may well be that more "intelligent persons do not gain more with practice, but that gain for such persons is relatively more permanent or, in the language of information processing, the results of such
gain are more easily and efficiently retrieved. Thus, achievement may measure cumulative gain from a number of situations, regardless of the absolute gain experienced at any particular time. Again in the terminology of information processing, more intelligent persons may not differ significantly in their input or immediate processing mechanisms, whereas, their storage mechanisms may be superior. In such a case, measures of achievement may be the best measures of "intelligence".

Another possible difficulty with the commonsense definition, which identifies intelligence as the ability to learn, arises insofar as both intelligence and learning ability must be considered general abilities according to the definition. Recent theories of intelligence developing from factor analytic studies (Guilford, 1959, 1966) have seriously questioned the general, unitary nature of intellectual ability. The generality of learning ability has been challenged by Woodrow (1946) in a review of findings from the experimental laboratory. The issues posed by Woodrow shall be returned to later.

We may now turn to the experimental evidence relating intelligence to learning ability.

The Negative Evidence

Hall (1936) posed the question of whether there is a "general learning ability which enables some individuals
to manifest considerable and rapid improvement in a wide range of learning situations and others to perform poorly on the same series of tasks? (p. 179)" In reviewing previous investigations relevant to the question, Hall found that the median of positive coefficients between learning tasks was .25. Only a third of the coefficients differed significantly from zero. Based only on these results, one is forced to conclude that a general learning ability, if it exists, is of only slight importance in determining learning performance. However, Hall criticized the methodology of these previous investigations on a number of grounds. Too often samples employed suffered from restriction in the range of talent. Studies using college students as subjects involve severe restrictions of range of talent due to pre-selection. Such a restriction of range of talent systematically lowers the correlation coefficient. Unreliability of measurement may also be a difficulty since the correlation between traits is restricted by the reliability of the measures employed. Also, Hall suggested that previous studies had suffered from limitations of number of subjects and that correlations should be based on at least 30 to 50 subjects. Low correlations between learning tasks may be attributable to the measures of learning used. Learning measured by absolute gain may differ from learning measured by total performance which resembles more achievement than actual learning, since total performance is affected by initial level of
performance. Other methodological considerations specified by Hall include the number of learning trials employed and the types of learning tasks employed.

Since most of the extraneous variables listed above have the effect of lowering the correlation coefficient, Hall reasoned that introduction of proper controls would raise the coefficients. He performed an experiment supposedly designed to correct for previous methodological errors. Four learning tasks of known reliability were employed, namely: a Stylus Maze; the Peterson Rational Learning Test; a List of Nonsense Syllables; and a Punchboard Maze. Having chosen four appropriate learning tasks according to his criterion of dissimilarity and known reliability, Hall proceeded to violate his own warning against limitation in range of talent. Subjects employed in the experiment were 100 college sophomore women volunteers. It is doubtful that he could have limited the range of talent much more if he had intentionally done so. The four tasks used were further so dissimilar as to press the limits of general learning ability, if such an ability exists. Subjects practiced each task once a week over a fourteen-week period. Intentional and unintentional extra-experimental practice would be difficult, if not impossible, to evaluate. Learning was measured by total errors, total number correct, and absolute gain. Reliability coefficients for total errors ranged in
the .90's; whereas, those for absolute gain ranged from .44 for stylus maze to .87 for nonsense syllables. Hall suggested that the high reliability of gains in learning nonsense syllables may be explained by the fact that it is the only task which excludes many chance factors such as differences in initial levels of performance. It is perhaps not surprising that none of the coefficients between gains differed significantly from zero. It is unfortunate that Hall was able to enumerate extraneous variables which should be controlled and then proceeded to violate his own warnings. The restriction in range of talent introduced by using women college sophomores could have easily been avoided by sampling from a high school or junior high school population. Although a general learning ability was certainly not demonstrated, methodological weaknesses in the experiment prevent conclusions concerning the extent to which it was discredited.

Husband (1939), in a similar experiment, correlated seventeen learning and memory tasks with IQ scores. Categories of the learning and memory tasks included: rote, motor and ideational. Correlations of the seventeen learning tasks with intelligence were so low as to be practically zero. Husband concluded that we should speak of learning abilities (plural) and not learning ability. However, it is noteworthy that the coefficients of correlation were slightly higher among complex functions than among simple motor and rote functions.
In order to investigate the possible effects of restrictions in range of talent upon obtained correlation coefficients, Husband (1941) chose junior high school students as subjects. Six learning tasks were employed including motor, rote, and ideational tasks. The median intercorrelation was found to be only .10 and Husband again concluded that learning abilities are specific. However, it is of interest to consider the nature of the tasks employed. Three of the tasks used were: code substitution; spool packing; and mirror drawing. None of these tasks correlated significantly with intelligence. However, the correlation between a reading comprehension task and intelligence test scores was .52. One may question the extent to which we should expect learning of extremely simple tasks to correlate with intelligence. Also, speed of learning simple motor tasks may not be expected to be related to intellectual ability, assuming that such an ability exists and our tests measure it. It may well be that more intelligent persons are able to learn more complex and challenging tasks which less intelligent persons are unable to master. For example, a severely retarded person is unable to master symbolic logic given any amount of practice, whereas the most intelligent of persons is presumably able to master such tasks. It is also possible that simple tasks are relatively non-challenging and non-motivating to persons possessing a high level of intellectual ability.
Such an hypothesis is suggested by Hall (1936).

Heese (1942) likewise investigated the possible existence of a general factor of improvement with practice. He presented six simple tests to 50 university students. A total of ten trials were administered on each test. Here again restriction of range of talent may be a factor since university students were used as subjects. Measures of gain were absolute gain and residual gain (defined as the difference between actual final scores and final scores predicted from initial scores). The six tests employed were: Addition, in which the subject summed three-digit numbers for a period of five minutes; Mirror Drawing; Stylus Maze; Sorting 20 two-digit numbers; Double Hand Test in which the subject manipulated two handles to draw a line; and Tapping, which required the subjects to make as many marks as possible on a piece of square paper with 1-inch rulings. Although improvement on mirror tracing correlated .57 with double-handle, most of the other measures of gain correlated positively, but low. A factor analysis employing the centroid method was applied to the data. Heese concluded that no general factor could be established.

Considering the nature and simplicity of the tasks employed, one would expect a simple factor of speed to contribute much to performance. This is precisely what was found. Speed of movement, memory and perception were names given to the three factors extracted from the matrix. With
such a restriction in the range of talent, a general ability which may exist would be expected to be held constant, thus allowing only relatively specific factors to come into play. Hull (1927) hypothesized that intra-individual abilities may closely follow the normal distribution found with inter-individual measures of ability. General ability, insofar as it exists, may be represented by the general elevation of the curve. Therefore, with a pre-selected group of college students, the general elevation of individual curves may be held constant, thus, revealing no general learning ability.

From an extensive review of the literature dealing with the ability to learn, Woodrow concluded that:

1. The ability to learn cannot be identified with the ability known as intelligence.

2. Individuals possess no such thing as a unitary learning ability.

3. Improvement with practice correlates importantly with group-factors, that is, relatively narrow abilities, and also with specific factors.

4. Even the group-factors involved in learning are not unique to learning, but consist of abilities which can be measured by tests given but once (Woodrow, 1946, pp. 148-149).

Since the studies reviewed by Woodrow contain many of the methodological problems previously mentioned such as limitations of range of talent and simplicity of the tasks employed, criticisms previously enumerated are relevant to Woodrow's conclusions.
In a study by Simrall (1947) the methodological difficulty of restriction in range of talent may have been avoided. Although no clear indication of the range of intelligence is given, the subjects used were 95 University High School students at the University of Illinois. Since the high-school population sampled may have been somewhat atypical and pre-selected, however, restrictions in range of intelligence may still have been a factor.

Simrall attempted to investigate the relation between intelligence as measured by the Otis Test and learning defined operationally as gains due to practice on a task presented repeatedly. The tasks used were a jumbled word test and a backwards writing test. Although mental age correlated .60 with initial score and .61 with final score on the backward writing test, and .59 with initial performance (.49 with final performance) on the jumbled words test; the correlations between mental age and gain with practice on the tasks were .08 and .28, respectively. Further, the correlation between gains due to practice of the two tests was only .18, indicating the absence of a general learning ability for the two tasks employed.

Simrall (1947) concluded that: "Every result of this experiment is contrary to the results predicted from the theory that intelligence is the ability to learn (p. 43)."

One apparent difficulty in Simrall's study is that no reliability coefficients for gain scores are given. Low reliability of gain scores may account for the negligible correlations between gain scores. Another possible source of
difficulty with absolute gain scores employed by Simrall is ceiling effects imposed by the task. Since the correlations which Simrall found between mental age and initial performance were high and positive, we may conclude that more intelligent persons began the task at a higher level of performance (i.e., closer to asymptote). Such subjects had, thus, less possible gain. Further, it may be hypothesized that more intelligent subjects gained most before performance commenced, especially on the backward writing test. Between instructions and the first performance trial, much could have been gained through covert practice. Moreover, gain at different levels of performance may have been qualitatively different. At advanced stages of performance on the backward writing task, for example, simple motor speed involved in recording responses may have been a factor of relatively greater importance than at lower levels of performance. Support for this hypothesis may be gathered from the finding of Fleishman (1960) and Fleishman and Hempel (1955) that different factors or abilities contribute to different stages of practice of a psychomotor task. Ferguson (1956) has hypothesized that "cognitive abilities play a more important role in the earlier stages of learning a motor task than in the later stages, when performance becomes organized in the form of a habitual psychomotor response pattern (p. 127)."

One study which found a negligible relation between "learning ability" and intelligence as measured by intelligence
tests is worth noting, primarily for its methodological weaknesses. Green (1953) presented five subtests of "learning ability" to 41 high school juniors. The subtests included: Letter Observation; Digit-Symbol; Vowel-Consonant; Parentheses Marking; and Reversal Type. Correlations between subtests ranged from -.12 to .58. The learning tests correlated .19 with intelligence as measured by the Otis Quick-Scoring Mental Ability Test and .16 with school grades; whereas, intelligence test scores correlated .54 with school grades. Green concluded that the "low correlations indicate that there is little relationship between the ability to learn and intelligence as measured by intelligence tests (p. 199)."

However, before accepting Green's conclusions at face value, it is advisable to first consider the procedure and results upon which they were based. The extent to which the five subtests measured learning is at least questionable. The usual method of measuring learning involves repeated presentation of a task from which a measure of gain, or the difference between initial and final scores, is taken. In Green's study, the five "learning subtests" were presented only once and the measures of learning were total scores. Further, the "learning subtests" were patterned after subtests from various intelligence tests and apparently differed little from these subtests. In fact, the five subtests taken as a whole rather resembled an intelligence test in that intercorrelations between subtests are low, but
generally positive, whereas, intercorrelations between the subtests and the test as a whole ranged between .48 and .79. Here the resemblance ceases. Whereas intelligence test scores correlated .54 with school grades and, thus, predict the criterion of school success fairly well, considering limitations of sample size (n = 41); the correlation between the total "learning" test and school grades was only .16. One begins to wonder whether Green devised five subtests of "learning" ability or whether he merely constructed a new type of intelligence test, and a poor one at that since it was unable to predict the criterion of school success.

Inspection of the subtests reveals that any learning involved was of a rather simple type. Without a measure of gain, it is questionable whether Green adequately measured any learning which may have been involved.

Another study which suffers from many of the same methodological weaknesses as those previously discussed was reported by Lindner and Overton (1960) who attempted to relate intelligence to grouping in learning. Forty-nine subjects from introductory psychology were used in the experiment. Such a sample suffers both from limited size and range of ability. The experimenters hypothesized that more intelligent persons easily discover the presumably helpful principle of grouping in learning and use it frequently. If this is so, one would expect most college sophomores to have learned the principle long before their second year of
college and to use it with about equal frequency. Since no data is given concerning the range of frequency of grouping, this hypothesis cannot be discarded. If everyone in the experiment groups material to be learned with approximately the same frequency, a high correlation between frequency of grouping and intelligence cannot be expected. It is perhaps not surprising that the correlation between frequency of grouping and intelligence was only .06.

A final methodological weakness arises from the fact that frequency of grouping was measured by the number of judges who said that the subject was grouping. Only three judges were used and presumably a Pearson product-moment correlation coefficient was used to compute the correlation between the frequency of subjects' grouping scores (maximum possible score per individual = 3) and intelligence scores. When the range of one of the variables is 0 through 3, the Pearson r is hardly an appropriate formula for computation of a correlation coefficient. To compound the error, no measure of the reliability of the judges' reports was presented. If perfect inter-judge reliability had existed, the grouping scores would have been dichotomous data since all three judges would have agreed that the subject was either grouping or he wasn't. Failure to demonstrate a relationship between intelligence and learning ability in this experiment was certainly not a guarantee that such a relation does not exist.
The last study to be reviewed which failed to find a positive relation between learning ability and intelligence was performed by Cautela (1965) who dealt specifically with probability learning. Again the range of talent was restricted to college students and the sample size was relatively small (n = 50). Subjects were presented with a two-choice probability learning situation in which one of the choices was reinforced 70% of the times. Subjects were given 200 trials and the last forty trials were used for purposes of the analysis. The correlation between the Otis scores and the probability scores was .15. Cautela concluded that no relation exists between intelligence and probability learning for college level subjects.

In the studies reviewed, thus far, which have found negligible relation between intelligence and learning ability, certain methodological weaknesses have been evident. A majority of studies suffered from restrictions in range of talent of subjects used since college students were often involved. Sampling from such a pre-selected population may have the effect of holding learning ability relatively constant, if such an ability indeed exists. Limitations in range of intelligence scores has likewise been evident. Sample size has also frequently been relatively small. When gain scores have been the measures of learning used, usually reliability coefficients have not been given. Finally, most of the learning tasks used have been rather simple and non-discriminating. If intelligence could be
demonstrated to predict learning in complex and challenging
tasks, limitations in prediction of simple tasks would do
little damage to our concept of intelligence or to our
confidence in using intelligence tests.

Even on simple tasks, the type of learning measure
used may be a critical variable. As Rapier has pointed
out:

Measurement by gain scores has the effect of
penalizing the bright students by not providing
enough ceiling for his growth. The more correct
responses the brighter student makes on his initial
score, the fewer test items he may have left to show
improvement on. Furthermore, the duller student may
be making his gain on answering the easier items
more correctly (Rapier, 1962, p. 8).

The Positive Evidence

One of the earliest studies which reported a positive
relationship between measured intelligence and learning
ability was reported by Johnson (1923). The learning task
involved was reading reversed print by holding a textbook
up to a mirror. One weakness of the experiment is that all
practice by the 60 university students was done outside
the laboratory during the subjects' free time. Subjects
were instructed to practice for ten minutes a day for
20 days. The correlation between the average score of
a number of group tests of intelligence and improvement as
measured by absolute gain in number of words read was
.46 ± .07. Johnson concluded that:
The results are, however, remarkable in showing such a close correlation when one considers the rather mechanical and uninteresting nature of the task of learning to read inverted print of a difficult thought content (Johnson, 1923, p. 544).

A possible difficulty in the experiment is the extra experimental practice required of subjects. Strict time limits and daily practice may not have been rigorously followed by all subjects.

Garrison (1928) investigated the relation between "rational" learning and intelligence. Tests of learning included the Rational Learning Test and Analogy Form Test devised by Peterson. Otis Intelligence Test scores correlated .50 ± .06 with the Analogy Form Test and .42 ± .06 with Rational Learning. Further, both of these tests slightly outperformed the Otis in predicting scores on weekly quizzes given by the experimenter.

Garrett (1928) presented eight memory-learning tests to a total of 158 college male freshmen. The tests included: Digit-span (auditory and visual); Paired-associates (auditory and visual); Logical memory; Digit-symbol; and Turkish-English. Taken individually, the tests had a low, but positive, correlation with the Thorndike Test of Intelligence. The median correlation was .21. However, when pooled, the eight tests correlated with the Thorndike .53, and .60 when corrected for attenuation. The latter correlation coefficient was as high as the correlation of the Thorndike Test with college grades. The similarity between the memory-learning tasks and subtests on standard intelligence tests is notable.
Also total scores on these tests were taken as measures of learning rather than gain scores.

An investigation of the relation between motor-learning and intelligence was performed by Spence and Townsend (1930). Only 20 subjects were used in the experiment which tested learning of a complex high relief finger maze. Subjects were selected on the basis of their scores on the Thurstone Intelligence Test. Subjects in one group were those who had achieved the highest scores on the intelligence test, while those in the other were those who had achieved the lowest scores. A large and highly significant difference was found between the two groups in their performance on the maze, with the group scoring highest on the Thurstone Test performing best on maze learning. The investigators present correlation coefficients with caution and for purposes of illustration only, since the full range of ability was not present in the sample and non-overlapping segments of the population were selected. Nonetheless, the correlations between trials, errors, time; and intelligence were respectively: .57 ± .11; .64 ± .09; .66 ± .08. The caution with which such coefficients must be interpreted, considering pre-selection and sample size, prohibits drawing conclusions.

Thompson and Witryol (1946) replicated the experiment of Spence and Townsend (1930) using 40 subjects whose intelligence test scores were continuously distributed. The measure of intelligence was obtained from the Otis Gamma Test, rather
than from the Thurstone Test used in the previous experiment. Although the intelligence tests scores of the sample used were continuously distributed, the range of scores was restricted by nature of the fact that subjects were all college students. The standard deviation of IQ's for the sample was only 8.2, whereas, the expected deviation for a representative sample should range between 15 and 16. The actual correlations between intelligence scores and errors and time were, respectively: .17 and .30. The investigators applied a statistical analysis to predict the correlations expected with a more representative and heterogeneous population. The expected coefficients of correlation between intelligence and trials, errors and time, respectively, were: .73, .74, and .76. Replication of this experiment with a representative population relative to dispersion of intelligence test scores, should produce interesting results.

In many of the experiments which have reported negligible relationships between learning ability and intelligence, gain scores are used as measures of learning. Tilton (1949) has focused criticism upon various measures of gain used. He emphasized that the reliability of the measures of gain should be as high as for end-scores. Even when coefficients of reliability for end tests are high, those for gain may be negligible. Many gain scores suffer from ceiling effects as previously discussed.
Tilton also recommended a rather novel approach to item analysis for tests measuring gain. Such an item analysis should not be based on per cent answering questions correctly, but on per cent learning to answer correctly.

Tilton investigated the effects of using improved measures of gain, according to his criterion. Gains were measured on achievement type tests given at fairly extensive intervals. For a seventh-grade group (N = 515) a test of twenty items was used which covered 11 weeks of instructions. For a twelfth-grade group (N = 134) a test of 54 items measured a full year's progress in social studies. For both tests, items were selected from longer tests by means of an item analysis of per cent learning. Results indicated that for the seventh grade, Otis IQ correlated .50 with initial scores and .49 with gains. For the twelfth-grade, Terman Group IQ scores were found to correlate .43 with initial performance and .49 with initial gain. When corrected for attenuation, correlations between IQ and gains were raised to .53 and .58.

Smith (1949) investigated the relation between intelligence test scores as measured by the California Test of Mental Maturity and learning resulting from the use of educational sound motion pictures. For three of four groups tested, the relationship was positive, significant and appreciable.

Gaudry and Champion (1962) presented a list of five paired associates for 16 trials to subjects divided into
high and low IQ groups. Latency of the correct response was the measure of learning performance. The high IQ group was equal to the low IQ group at the outset, and had the same upper limit of performance, but approached the upper limit of performance at a significantly faster rate. It was speculated that with more complex and challenging tasks, the upper limits as well as rate of approaching the limit would differ significantly for more intelligent groups.

As previously speculated, restriction in the range of talent, which occurs when college students are the subjects, may have the effect of holding "general learning ability" constant across subjects. Two studies reviewed (Tilton, 1953, and Mackay and Vernon, 1963) found considerable evidence for a general factor of learning ability among relatively non-selected grade school subjects. The study by Mackay and Vernon (1963) used residual gain as a measure of learning ability. In Tilton's study, absolute gain in achievement type tests was the measure of learning used.

Duncanson (1966, 1967) has investigated the interrelations between learning and ability measures and between learning in different situations by administering a battery of ability tests and nine learning tasks to 102 sixth-graders. The types of learning were concept formation, paired-associates and rote-memory tasks. Every learning task with the exception of concept formation, was found to be related to one or more of the ability tests. Tasks involving words showed highest loading on the verbal factor. Those involving
numbers loaded heaviest on the numerical factor, and those involving rote-memory tasks had highest loadings on the memory factor. On the other hand, three learning factors were found which were independent of abilities measured. Although no general learning factor was found, group learning factors were found.

McGeoch has hypothesized that:

A pool of a large number of measures of learning would correlate at least as highly with intelligence tests as intelligence tests do with each other (McGeoch, 1942, p. 251).

Harootunian investigated McGeoch’s hypothesis. A total of eight learning tasks and two intelligence tests (the CTMM and the Otis Beta) were administered to 88 eighth-graders. The median correlation between any task and intelligence scores was .36. The median correlation between learning tasks was .28. However, a composite learning score correlated .71 with the CTMM and .73 with the Otis Beta. Thus, when learning tasks were combined, the relationship between the learning tasks and the intelligence tests was almost the same as the correlation between the intelligence tests. Harootunian concluded that: "Intelligence and learning ability have much in common when the latter is measured by the composite of a number of scores (Harootunian, 1966, p. 213)."

**Programmed Learning**

As previously discussed, intelligence tests may be described as instruments which sample present achievement or
performance in a number of situations. The assumption underlying such instruments is that present level of performance (i.e., achievement) reflects previous learning or gain with practice. It is reasoned that amount of previous gain is a measure of ability to gain (Woodrow, 1946). As Sorenson pointed out:

When he administers an intelligence test, the psychologist samples the knowledge and skills which the subject has learned in the past, sometimes years ago. Neither the psychologist nor the subject can know very much about the conditions under which the learning occurred, nor what specific experiences were significant, nor how much time was involved. In any society, some subjects will have had opportunities and experiences not shared by others, because of differences in social class, etc. (Sorenson, 1963, p. 326).

Sorenson (1963) has suggested that cumulative records from "teaching machines" or programmed instruction could be used as an alternative to intelligence tests. He commented that intelligence may be defined as the "ability to learn from experience" (p. 325) and that "teaching machines" could potentially assess rate of learning in a way which is more meaningful and applicable than has been previously possible.

One type of program for instruction, the branching program, is illustrated. The student is presented with a unit of instruction on microfilm. After he studies that unit, a multiple choice question is presented on the screen and he selects one of the alternatives by depressing one of several buttons before him. If the question is answered correctly, the student is informed as such by the flashing
of a green light. If he has answered incorrectly, the material is presented again, his error is explained, and the student is again tested on an alternate question. The variety of types of instructional materials are potentially limitless. Advantages of this type of instruction are that the student is tested immediately after studying each unit and that the system includes feedback control. Also, for the purpose of measuring progress in learning, such machines can print out a cumulative record of the student's sequence of the choices, amount of time spent and number of attempts on a problem. Such a learning record would tell a great deal about the conditions under which the subject learned or failed to learn and a more precise and extensive record of progress may be gathered. Sorensen further pointed out that the problem of developing a "culture free" test would practically disappear.

One of the most promising prospects of using cumulative records from programmed instruction is that such measures may have greater predictive validity, especially when the criterion is academic performance. Further, since cumulative records from programmed instructions would be based upon larger samples of behavior than IQ scores, they may be expected to have greater reliability. In fact, a measure of learning ability could be based upon full year's cumulative record or greater, since machines could have the capacity to compute, store, and recompute at specified intervals any representative statistic desired. A child's
"learning ability" so measured could be compared with his appropriate referent population and with populations with which he is likely to compete.

Also programmed learning would greatly facilitate research into the general or specific nature of learning ability since extended records gained from "teaching machines" would probably produce measures of gains having greater reliabilities. The generally limited reliability of measures of gain has been a real methodological stumbling block in many previous investigations.

One possible problem with Sorenson's approach is the assumption that intelligence is best identified with "learning ability" as measured by simple rate of gain. As previously noted, long-term storage of gains made may also be an important component of intelligence, as may be the upper limit of gain in learning complex materials. However, both of these possibilities could be investigated by using programmed instruction devices. The main benefits of such devices would be the standardization possible in the presentation of materials and recording of responses. Also, coupled with computers of sufficient size, teaching devices could be made to more easily yield summary type statistical values by programming the computer to accept running records and perform desired statistical analysis on the data.

Wardrop and Dubois (1965) investigated the use of programmed instruction as a miniature learning situation for predicting performance in a classroom situation.
Residual gain was used as a measure of learning and subjects were 330 Air Force trainees. The General Classification subtest of the Basic Test Battery, a group test of verbal intelligence, was the measure of intelligence used; two tests of perceptual-motor skills were used. The programmed instruction consisted of an 85-minute linear program on study skills. Subjects were given a pre- and post-instruction test. Programmed instruction was found to be more closely related to classroom learning than the other learning tests, although none of the correlations exceeded .30. The intelligence measure used correlated with the criterion of classroom success only .35 and when intelligence measures were combined with programmed learning results, the correlation coefficient was raised only slightly to .38.

Although no definite relationships were demonstrated with such low coefficients of correlation, several things should be noted. The sample used was pre-selected and the programmed instruction was not optimal according to Sorenson's (1963) description of programmed instruction techniques.

**General Criticisms**

Several methodological difficulties have persisted in the investigations reviewed, such as restrictions in the range of talent; lack of agreement concerning appropriate measures of gain; lack of evidence for reliabilities of measures of gain; presence of ceiling effects in the tasks used which have the effect of inhibiting subjects who begin
at a higher level of initial performance; and at times, perhaps faulty statistical analysis.

When college sophomores are the subjects in investigations, a double restriction in the range of talent most probably occurs. One selection process occurs at the point of admission to college; another selection process is involved in the attrition rate through the first year. Thus, any sample of college sophomores probably has an extremely limited range of IQ scores. The reliability of IQ scores sampled from a college sophomore population is most probably appreciably less than for the general population. If such restriction in the range of ability is the case, one would not expect a high correlation between IQ scores of college sophomores and any other variable measured.

It has, likewise, been suggested that extremely simple tasks which have frequently been employed in investigations may not afford an adequate test of the relation between intelligence and learning ability. Such tasks frequently suffer from ceiling effects. Subjects who begin at a higher level of performance may have less potential gain remaining. Further, although gain on simple psychomotor type tasks may not be related to measures of intelligence, gain on complex verbal type tasks may. Few of us may be disturbed to discover that more "intelligent" persons are not able to learn backward alphabet printing or mirror star tracing faster than less intelligent persons.
However, if such persons do not gain more from reading complex verbal materials or from instruction in symbolic logic, it may well tell us something about the usefulness of our measures of intelligence.

The Present Investigation

The purpose of the present study is to investigate the relationship between gain in performance on verbal comprehension tests when repeated exposure is given to information type material, and measures of verbal and general intellectual ability.

Certain objectives were established at the outset in accordance with criticisms of previous research in the area. These objectives included:

1. The sample should be drawn from as heterogeneous population relative to the range of ability as possible.
2. The learning task should be sufficiently complex to avoid ceiling effects.
3. Several measures of gain should be computed and compared including absolute gain, proportion of possible gain and residual gain.
4. A measure of reliability of gain should be obtained.
5. A measure of relatively permanent gain should be obtained in order to determine the extent to which it differs from immediate gain, and its relationship to measured intelligence.
CHAPTER II

THE METHOD

Pretest of the Items

Subjects  A total of 96 eighth-grade students at a public junior high-school in the Richmond metropolitan area served as Ss for the pretest of the items. All Ss were volunteers from general classes. Special grouping of students into academic and vocational curricula does not occur until the ninth grade in this particular school system.

Materials All four alternate forms (i.e., Forms: Am (Revised); Bm (Revised); Cm and Dm) of the Rate and Comprehension: Part A subtests from the Advanced Test of the Iowa Silent Reading Tests, New Edition (Greene, Jorgensen & Kelly, 1943) were used as the basic materials from which the pretest was developed. The four comprehension passages deal with general science content and cover the topics: "Glass, Rubber, Cork, and Iron." The Iowa Test Booklet contains a total of ten 3-choice questions for each alternate subtest.
The manual for the test fails to give predictive or concurrent validity coefficients. However, evidence is presented for content validity (Greene, Jorgensen & Kelly, 1943, p. 3). For the total 1942 national standardization population of ninth-graders, the reliability coefficient computed by the Kuder Richardson Formula 21 was .728 (Greene, Jorgensen & Kelly, 1943, p. 5).¹

Procedure--A pretest of listening comprehension was developed by the experimenter using the basic materials from the four comprehension subtests of the Advanced Test of Iowa Silent Reading Tests. The four alternate comprehension passages were read on audio-tape by a semi-professional announcer. The passages were recorded in continuous fashion, each preceded by its appropriate title. The audio-tape record of the comprehension material shall hereafter be referred to as the "listening comprehension material."

In addition to the 40-test questions obtained from the Iowa subtests, E constructed an additional 75 questions which were patterned after the format of the Iowa subtest questions. Fortunately, each comprehension passage of the Iowa Silent Reading Tests is repeated in the Directed Reading subtest of an alternate form of the test. Therefore,

for each subtest, a number of alternate questions could be derived or adapted from the appropriate Directed Reading subtest.

The instructions to the Ss and the 115 questions were recorded by E on audio-tape. Each question number was announced, followed by the question and the three possible alternatives. A ten-second delay preceded the reading of the next complete question.

The question numbers and possible alternative answers were printed on the pretest form with a space in the margin in which S could record the letter of the answer which he considered the correct one.

In the actual pretest situation, all Ss were tested in a group. Testing commenced at 8:45 A.M. and lasted for a duration of approximately 1 hr. 30 min. After 45 min., a short break was given in the testing situation. The testing session was conducted in the school cafeteria.

At the outset of the pretest, Ss were presented with the answer sheets which were placed face down in front of them. Ss were instructed to print their name, age, and homeroom number on the back of the answer sheets.

The tape was then begun with instructions to Ss which were as follows:

You will shortly hear a tape which covers four topics. Listen carefully to the tape. When it is finished, you will be tested in order to see how much you have learned. Are there any questions? (pause)
Can everyone hear me? (pause)

After a delay for questions, the taped "listening comprehension material" was presented. Output came from the speaker systems of two tape recorders in order to assure sufficient level of volume and clarity. At the conclusion of the "listening comprehension material," the following instructions were presented on tape in order to introduce the pretest.

Please do not turn over the answer sheets until you are told to do so. You will hear questions which you are to answer on the answer sheets. Following each question, three possible answers will be given. These three possible answers are written on your test sheet next to the question number. After each question has been read and the three possible answers have been given, you are to place the letter of the answer which you think is the correct one in the margin next to the question number. Please wait until the question has been read and all the possible answers have been given before you record your answer. However, DO NOT GUESS. A portion of the number which you get wrong will be subtracted from the number which you answer correctly. Therefore, guessing may actually lower your score. If you do not know the answer to a question, leave it blank. You will have ten seconds in which to answer each question. Are there any questions? (pause)

Now turn your answer sheets over. (pause)
Listen to the following sample question: 'What is the basis of sand?' . . . Now, look at the sample answers. The sample answers are: (a) soda; (b) silica; (c) lead. The correct answer is (b) silica. Before the test begins, are there any questions? (pause)

The test will now begin. Remember, do not guess. If you do not know the answer to a question, leave it blank. Work only on the question which is announced and wait until the question is read and the three possible answers have been given before you answer the question.
Item Analysis—An item analysis was then performed on the pretest items. An index of item difficulty, $P$ (defined as the proportion of $S$s passing the item), was computed for each item as a rough index of item difficulty. The point-biserial correlation coefficient between each item and the test as a whole was estimated by using an abac presented in Guilford (1954, p. 429). The decision to employ the point-biserial correlation coefficient was based upon Guilford's suggestion that "the point-biserial $r$ is the most appropriate coefficient of correlation to use for a realistic indication of item-criterion correlation" (1954, p. 432)."

The items retained from the pretest were those for which the point-biserial $r$ with the entire test equalled or exceeded the critical value of $.205$ ($df = 95$) for significance at the .05 level. A total of 75 items were retained in accordance with the above criterion.

Three alternate forms of a "listening comprehension test" were constructed following a scatter diagram technique outlined in Guilford (1954, p. 443). (The proportion passing each item and the point-biserial $r$ of each item with the test are given in Table A of Appendix I.)

The means, mean item-test correlations and estimates of internal-consistency reliabilities of the three alternate forms of the "Listening Comprehension Test" based on the item pretest sample, are presented in Table I.
### Table I

Means, mean item test r's, and internal-consistency reliabilities of the three alternate forms of the "Listening Comprehension Tests" for the Item Pretest Sample. \(^1\)

<table>
<thead>
<tr>
<th>Alternate Test Form</th>
<th>Mean Item-Test r</th>
<th>Internal-Consistency Reliabilities (^2)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>12.26 .35</td>
<td>.78</td>
</tr>
<tr>
<td>2</td>
<td>12.27 .34</td>
<td>.76</td>
</tr>
<tr>
<td>3</td>
<td>12.27 .33</td>
<td>.75</td>
</tr>
</tbody>
</table>

\(^1\) Tests were scored without correction for guessing.

\(^2\) Computed by Spearman-Brown formula following a procedure specified by Guilford (1965, p. 463).
Subjects: A total of 83 eighth-grade students, from three homeroom classes, currently attending a public junior high school in the Richmond metropolitan area, served as Ss in the experiment. However, due to the absence of ability scores for five Ss, they were dropped from the analysis. The means and standard deviations of various ability test scores for the sample are presented in Table II.

Materials: Ability tests scores were obtained from the files of the school system. These measures included two scores from the Differential Aptitude Test, Form L, and two scores from the California Test of Mental Maturity, Form JH. The Differential Aptitude Tests had been administered to the Ss in October, 1967, approximately six months prior to the experiment. The California Test of Mental Maturity had been administered in October, 1966 (approximately 18 months before the experiment was conducted). The two scores taken from the DAT were Verbal Reasoning subtest (VR), and a combined score for Verbal Reasoning and Numerical Aptitude subtests (VR/NA).

After a careful review of 4,096 validity coefficients presented in the manual for the DAT, McNemar (1964) concluded that VR is the best single predictor. Verbal Reasoning is essentially a test of analogies. When combined in unweighted fashion with Numerical Aptitude, the composite score "serves the same purpose as the group tests of general ability . . . (Cronbach, 1960, p. 271)."
<table>
<thead>
<tr>
<th>Test</th>
<th>Mean</th>
<th>Standard Deviation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Differential Aptitude Test:</td>
<td></td>
<td></td>
</tr>
<tr>
<td>VR (Verbal Reasoning)</td>
<td>20.71</td>
<td>9.03</td>
</tr>
<tr>
<td>VR ≠ NA</td>
<td>38.18</td>
<td>12.96</td>
</tr>
<tr>
<td>California Test of Mental Maturity</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Verbal IQ</td>
<td>108.23</td>
<td>12.95</td>
</tr>
<tr>
<td>Total IQ</td>
<td>105.80</td>
<td>11.37</td>
</tr>
</tbody>
</table>
The test manual gives the predictive validity of the DAT VR/NA with four-year high school grades as .73 for boys and .71 for girls. For a sample of high-school seniors, the estimate of construct validity with the CERB-SAT-V was .70 for boys and .72 for girls. Construct validity with the Wechsler Adult Intelligence Scale for ages 16-17 was given as .79 (Psychological Corporation, 1958). Table III presents the means and standard deviations of the DAT VR and DAT VR/NA for the national sample of eighth-grade pupils.

According to the manual for the 1963 Revision, the California Test of Mental Maturity was originally designed as a group test of intelligence patterned after the individual Stanford-Binet. According to the test manual, the correlation between the language IQ scores and high school grades in science was .60, while the total IQ scores correlated .57 with grades in science. The coefficient of construct validity between the CTMM total IQ scores and the Wechsler Intelligence Scale for Children was given as .77 (Cronbach, 1960, p. 116).

### TABLE III
AREAS AND STANDARD DEVIATIONS OF DIFFERENTIAL
APTITUDE TEST SCORES FOR THE EIGHTH GRADE
STUDENTS OF THE NATIONAL SAMPLE*

<table>
<thead>
<tr>
<th>Test</th>
<th>Mean Boys</th>
<th>Mean Girls</th>
<th>Standard Deviation Boys</th>
<th>Standard Deviation Girls</th>
</tr>
</thead>
<tbody>
<tr>
<td>DAT VR</td>
<td>15.7</td>
<td>16.0</td>
<td>8.3</td>
<td>8.0</td>
</tr>
<tr>
<td>DAT VR / NA</td>
<td>29.5</td>
<td>30.5</td>
<td>13.6</td>
<td>13.1</td>
</tr>
</tbody>
</table>

The current revision of the CTMM utilizes deviation IQ's with a mean of 100 and a standard deviation of 16 IQ points for each age group.3

Procedure. All subjects were tested in their natural class groups and in a classroom situation. Because all materials were presented in the form of tape-recorded instructions and mimeographed tests, standardization of test administration was maximized.

The most salient points of the procedure will first be presented as an over-view. Ss were administered the first "Listening Comprehension Test" prior to formal exposure to the "listening comprehension material." This "Pretest," as it shall hereafter be referred to, was given for purposes of obtaining a measure of the amount of relevant knowledge which Ss had prior to the presentation of the "comprehension material." In other words, the "Pretest" was designed to afford a measure of an initial base rate prior to the actual learning experience. The "listening comprehension material" was then presented, followed by the second alternate form of the "Listening Comprehension Test," which shall hereafter be referred to as "Learning Test I." A second

presentation of the tape was followed by "Learning Test II" (the third alternate form of the "Listening Comprehension Test").

After a duration of 24 hours, a "Retention Test" (Learning Test II repeated) was administered. The specific details of the procedure were as follows:

At the outset of the experiment, three answer sheets were placed face down on each S's desk. Ss were asked to print their name, class, and age on the back of the answer sheet. The tape was then begun with the following instructions:

Please do not turn over the answer sheets until you are told to do so. You will shortly hear 25 questions which you are to answer on the top answer sheet. Following each question, three possible answers will be given. These possible answers are printed on your answer sheet next to the question number.

After each question has been read and the three possible answers have been given, you are to put the letter of the answer which you think is the correct one in the space provided for it next to the question number. However, DO NOT GUESS! A portion of the number which you get wrong will be subtracted from the number which you answer correctly. Therefore, guessing may actually lower your score. If you do not know the answer to a question, leave it blank.

Also, work only on the question which is announced, do not go back and answer questions which you have missed or which you think you have answered incorrectly.

When working on a question, wait until the question has been read and the three possible answers to the question have been given before you answer the question.
You will have ten seconds in which to answer each question. Are there any questions? (pause)

Now, turn over your answer booklets (pause). Notice the words "Sample Question" on the front of the first page. Listen to the sample question. 'What raw material is most often used in the process of making glass?' Notice that the possible answers given are: (a) rock; (b) iron; (c) sand. The correct answer is (c) sand. Place the letter (c) in the space provided for it next to the question number (pause).

Do not worry if you do not know the answers to the questions. You will shortly have an opportunity to learn about the information which the tests cover. Before the test begins, are there any questions (pause)?

The test will now begin. Remember, do not guess. Work only on the question which is announced and wait until the question has been read and all of the possible answers have been given before you answer the question.

Now open your answer booklets to the first test. At the top of the page, the words 'Test 2' should be printed. Does everyone have Test 2 (pause)?

The test proceeded. Each question number was announced. The question was then read followed by the possible alternatives. Before the beginning of the next question, a 10-second interval was allowed.

Following the Pretest, instructions were given concerning the "comprehension material."

You will now hear passages about glass, rubber, cork, and iron. Listen carefully. After you have heard these passages, you will be tested to see how much you have learned. Are there any questions (pause)?

The taped "comprehension material" was then presented followed by "Learning Test I." The following instructions were presented for "Learning Test I":
Now, you will be tested to see how much you have learned. Each question will be given as before. Remember, DO NOT GUESS! Work only on the question which is announced and wait until the question has been read and the three possible answers have been given before you answer the question. Turn to the second test. At the top 'Test Y' should be printed. Does everyone have 'Test Y' (pause)?

Are there any questions?

A 5-minute rest followed "Learning Test I" during which Ss were permitted to be excused from the room.

A second presentation of the tape was preceded by the following instruction.

You will again hear the three passages. Listen carefully. When the tape is finished, you will be tested to see how much you have learned. Are there any questions?

Immediately following the second presentation of the tape, "Learning Test II" was administered, preceded by the following instructions:

You will now be tested to see how much you have learned. Remember, DO NOT GUESS! Work only on the question which is announced and wait until the question and all possible answers have been given before you answer the question. Now turn to the third test. At the top of the test, the words 'Test X' should appear. Does everyone have 'Test X'?

Are there any questions?

"Learning Test II" proceeded as the previous learning test. At the conclusion of "Learning Test II," the following announcement was presented on the tape:

The test is now complete. Thank you for your cooperation. Please do not discuss the test with each other or with your friends in other classes.
After a 24-hour interval, Ss were administered the "Retention Test" ("Learning Test II repeated"). The following instructions preceded the test:

You will be tested as before on the material which you heard. Remember, **DO NOT GUESS!** Work only on the question which is announced and wait until the question is read and all possible answers are given before you answer the question.

At the conclusion of the "Retention Test," the Ss were again requested not to discuss the test.

A transcript of the three alternate forms of the "Listening Comprehension Test" may be found in Appendix II. With each test is an example of the answer sheet for that test.

**Analysis of the Data.** With the exception of a few minor computations, the analysis of the data, including scoring of the tests, was accomplished with the aid of an I.B.M. 1620 computer. Program number 6.0.148, a "Single and Multiple Linear Regression Analysis Program" from the 1620 General Program Library, was employed in the analysis. A brief description of the program may be found in Appendix III-A. All other programs employed in the analysis of the data were written by E in Fortran IV. A listing of the titles of the programs written by E may be found in Appendix II-B.

**Scoring of Items.** A crucial variable in experiments measuring gain in performance is the relative difficulty of items measuring gain. Unless all items are equal in difficulty, gain may become more difficult at higher levels of
performance, since presumably only more difficult items remain on which to make gain. If all items are given a score of 1 or 0, persons beginning at higher levels of performance may be differentially disadvantaged by having to make gains on more difficult items. In the present investigation, there was a potential built-in control for differential item difficulty. The control was derived from the pretest of the items. The number of Ss in the item pretest sample who answered an item incorrectly (here, no response is considered an incorrect response) was taken as a measure of the relative difficulty of each item. In scoring the "Listening Comprehension Tests," items were weighted by the number of Ss who answered the item incorrectly in the item pretest sample.

In scoring the tests, the conventional formula for correction for guessing, "Right minus \( \frac{\text{Wrong}}{n-1} \)" (Cronbach, 1960, p. 50), was employed. In the context of the present investigation, the formula becomes "Weighted Rights minus \( \frac{\text{Weighted Wrongs}}{n-1} \)."

**Measures of Gain**

Several measures of gain were employed in computing gain in performance on the "Listening Comprehension Tests." These include:

1. **Absolute gain** Defined as the arithmetic difference between final performance and initial performance. Absolute gain is probably the least sensitive of the measures
of gain employed since it does not take into account differences in initial status.

2. **Proportion of possible gain** defined as the ratio of absolute gain to possible gain (maximum possible score minus initial score). Proportion of possible gain should compensate for any ceiling effects present by affording a statistical advantage to Ss who begin at a higher level of performance. Such individuals presumably have less remaining potential gain.

3. **Residual gain** defined as that portion of the measure of final performance which is statistically independent of initial status (Wardrop and DuBois, 1965). Residual gain is computed as the difference between the final score which is achieved and the final score predicted from actual initial score. In the computation of residual gain, a regression equation based on the correlation between initial and final scores is derived. From the equation, each S's predicted final score is computed from his initial score. His residual gain is then the difference between his predicted final and actual final scores, as previously stated. Residual gain is perhaps the most sensitive of the gain measures. Wardrop and DuBois state that "it (residual gain) offers the advantages of consistency, adaptability, and statistical logic..." (1965, p. 327)."
CHAPTER III

THE RESULTS

The means and standard deviations of the weighted "Listening Comprehension Tests" (LCT) Scores are presented in Table IV. The mean weighted scores of the "Listening Comprehension Tests" are presented graphically in Figure 1. Results of a single classification analysis of variance for repeated measures (see Table V) revealed a significant difference between the mean weighted "LCT" scores \( F(4,76) = 95.32, P < .01 \). A Duncan analysis revealed that the mean score of the "Pretest" was significantly \( (P < .01) \) below the means of the other three "Listening Comprehension Tests." However, the differences between the mean scores of the other three tests (i.e., "Learning Test I," "Learning Test II," "Retention Test") did not reach statistical significance \( (P > .05) \). The preceding results of the analysis of variance and the Duncan test may be interpreted with caution in that mean scores are based on alternate test forms.
TABLE IV
THE MEANS AND STANDARD DEVIATIONS OF THE WEIGHTED "LISTENING COMPREHENSION TESTS" SCORES

<table>
<thead>
<tr>
<th>Test</th>
<th>Mean</th>
<th>Standard Deviation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pretest</td>
<td>55.05</td>
<td>150.70</td>
</tr>
<tr>
<td>Learning Test I</td>
<td>433.33</td>
<td>245.63</td>
</tr>
<tr>
<td>Learning Test II</td>
<td>464.74</td>
<td>291.91</td>
</tr>
<tr>
<td>Retention Test</td>
<td>473.73</td>
<td>292.76</td>
</tr>
</tbody>
</table>
Figure 1. Mean Weighted Scores on the "Listening Comprehension Tests"
### TABLE V

SINGLE CLASSIFICATION ANALYSIS OF VARIANCE (REPEATED MEASURES) BETWEEN LOG TRANSFORMED WEIGHTED SCORES ON THE "LISTENING COMPREHENSION TESTS"

<table>
<thead>
<tr>
<th>Source</th>
<th>S.S.</th>
<th>d.f.</th>
<th>M.S.</th>
<th>F</th>
</tr>
</thead>
<tbody>
<tr>
<td>Between people</td>
<td>3.1</td>
<td>76</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Within people</td>
<td>5.6</td>
<td>231</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Tests</td>
<td>3.14</td>
<td>3</td>
<td>1.04</td>
<td>95.32**</td>
</tr>
<tr>
<td>Residual</td>
<td>2.5</td>
<td>228</td>
<td>.011</td>
<td></td>
</tr>
</tbody>
</table>

** F.99 (3,228) = 3.88

Scores were transformed to logs to the base 10 in order to satisfy the assumption of homogeneity of variance.
### TABLE VI

DUNCAN PROCEDURE FOR TESTING DIFFERENCES BETWEEN LOG TRANSFORMED WEIGHTED TOTAL SCORES ON THE "LISTENING COMPREHENSION TEST"

<table>
<thead>
<tr>
<th></th>
<th>Pretest</th>
<th>Learning Test I</th>
<th>Learning Test II</th>
<th>Retention</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Ordered Totals</strong></td>
<td>(1)</td>
<td>(2)</td>
<td>(3)</td>
<td>(4)</td>
</tr>
<tr>
<td></td>
<td>210.03</td>
<td>227.45</td>
<td>228.06</td>
<td>228.41</td>
</tr>
<tr>
<td><strong>Critical Values (5%):</strong></td>
<td>2.60</td>
<td>2.74</td>
<td>2.83</td>
<td></td>
</tr>
<tr>
<td><strong>(1%):</strong></td>
<td>3.46</td>
<td>3.61</td>
<td>3.71</td>
<td></td>
</tr>
<tr>
<td><strong>Ordered Differences</strong></td>
<td>(2)</td>
<td>(3)</td>
<td>(4)</td>
<td></td>
</tr>
<tr>
<td>(1)</td>
<td>*17.43</td>
<td>*18.03</td>
<td>*18.38</td>
<td></td>
</tr>
<tr>
<td>(2)</td>
<td>.61</td>
<td>.96</td>
<td>.35</td>
<td></td>
</tr>
<tr>
<td>(3)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

* $P < .05$
<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>CTM-L</td>
<td>Language IQ score from the California Test of Mental Maturity.</td>
</tr>
<tr>
<td>CTM-TOT</td>
<td>Total IQ score from the California Test of Mental Maturity.</td>
</tr>
<tr>
<td>DAT-VR</td>
<td>Verbal Reasoning subtest score from the Differential Aptitude Tests.</td>
</tr>
<tr>
<td>DAT-VR / NA</td>
<td>Score composed of unweighted combination of Verbal Reasoning and Numeric Ability subtest scores of the Differential Aptitude Tests.</td>
</tr>
<tr>
<td>PT</td>
<td>&quot;Pretest&quot; score (alternate form number 1 of the &quot;Listening Comprehension Test&quot;).</td>
</tr>
<tr>
<td>LTI</td>
<td>Score on &quot;Learning Test I&quot; (alternate form number 2 of the &quot;Listening Comprehension Test&quot;).</td>
</tr>
<tr>
<td>LTII</td>
<td>Score on &quot;Learning Test II&quot; (alternate form number 3 of the &quot;Listening Comprehension Test&quot;).</td>
</tr>
<tr>
<td>RT</td>
<td>&quot;Retention Test&quot; score (alternate form number 3 of the &quot;Listening Comprehension Test&quot; repeated).</td>
</tr>
<tr>
<td>Gain 1</td>
<td>Gain in performance on the &quot;Listening Comprehension Test&quot; which occurred with the first presentation of the &quot;comprehension material&quot; (i.e., LTI - PT).</td>
</tr>
<tr>
<td>Gain 2</td>
<td>Gain in performance on the &quot;Listening Comprehension Test&quot; which occurred with the second presentation of the &quot;comprehension material&quot; (i.e., LTII - LTI).</td>
</tr>
</tbody>
</table>

(cont'd.)
| Gain 1 & 2 | Total gain in performance on the "Listening Comprehension Test" with both presentations of the "comprehension material" (i.e., LTII - PT). |
| Gain 4    | Amount of total gain which was retained over a 24-hour period (i.e., RT - FT). |
| Gain 5    | Amount of Gain 2 which was retained over a 24-hour period (i.e., RT - LTII). |
Coefficients of Correlation Between Test Scores

Table VIII presents the coefficients of correlation between the weighted raw test scores on the "Listening Comprehension Tests," and between these scores and measures of ability. All tests of statistical significance in this and subsequent sections of the analysis were performed at the .05 level of significance.

Several correlation coefficients should be noted. First, none of the coefficients of correlation between the scores on the "Pretest" and the other "Listening Comprehension Tests" (LCT) differed significantly from .00. Further, none of the \( r \)'s between the "Pretest" and the measures of ability were statistically significant.

However, the remaining LCT test scores correlated significantly with each other and with measures of ability. For example, LTI ("Learning Test I") correlated .67 with LTII. Similarly, the coefficient of correlation between LTI scores and scores on the DAT-VR was .59. This latter measure may be taken as an estimate of the construct validity of LTI with the DAT-VR, for the sample employed. Other estimates of construct validity of LTI were: .49 with the CTMM-L, and .61 with the DAT-VR/NA. For "Learning Test II," estimates of construct validity were: .60 with the DAT-VR; .51 with the CTMM-L; and .62 with the DAT-VR/NA. For the "Retention Test" (Learning Test II repeated after a 24-hour interval), the estimates of construct validity were: .48 with the DAT-VR; .44 with the CTMM-L; and .46 with the DAT-VR/NA.
TABLE VIII

7 X 7 INTERCORRELATION MATRIX FOR RAW TEST SCORES (N = 77)

<table>
<thead>
<tr>
<th></th>
<th>PT (1)</th>
<th>LTI (2)</th>
<th>LTII (3)</th>
<th>RT (4)</th>
<th>DAT-VR (5)</th>
<th>CTMM-L (6)</th>
<th>DAT-VR # NA (7)</th>
</tr>
</thead>
<tbody>
<tr>
<td>PT (1)</td>
<td>1.00*</td>
<td>.09</td>
<td>.17</td>
<td>.12</td>
<td>-.10</td>
<td>-.17</td>
<td>.03</td>
</tr>
<tr>
<td>LTI (2)</td>
<td>1.00*</td>
<td>.67*</td>
<td>.61*</td>
<td>.59*</td>
<td>.47*</td>
<td>-.10</td>
<td>.61*</td>
</tr>
<tr>
<td>LTII (3)</td>
<td>1.00*</td>
<td></td>
<td>.84*</td>
<td>.60*</td>
<td>.51*</td>
<td>.47*</td>
<td>.62*</td>
</tr>
<tr>
<td>RT (4)</td>
<td>1.00*</td>
<td></td>
<td></td>
<td>.48*</td>
<td>.44*</td>
<td>.48*</td>
<td></td>
</tr>
<tr>
<td>DAT-VR (5)</td>
<td>1.00*</td>
<td></td>
<td></td>
<td></td>
<td>.75*</td>
<td>.73*</td>
<td></td>
</tr>
<tr>
<td>CTMM-L (6)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1.00*</td>
<td></td>
<td></td>
</tr>
<tr>
<td>DAT-VR # NA (7)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1.00*</td>
<td></td>
</tr>
</tbody>
</table>

1 See Table VII for listing of abbreviations.

* P < .05
For the sample employed, the estimated construct validities of the CTMM-L were .75 with the DAT-VR, and .73 with the DAT-VR/NA. In addition, the estimate of construct validity of the CTMM-TOT with the DAT-VR/NA, for the sample tested, was .60 (see Table XI).

The test-retest reliability of "Learning Test II" may be estimated by the correlation between scores on the first administration of the test and scores on the test administered after a 24-hour delay (otherwise termed the "Retention Test"). The coefficient of correlation was .84.

**Coefficients of Correlation between the Measures of Gain**

Table IX presents the intercorrelations between two measures of gain for the three methods of computing gain. It should be noted that Subject #48 had been dropped in computing Proportion of Possible Gain because the computation of that measure of gain was impossible for his raw scores due to the occurrence of a zero in the denominator of the computational formula. For purposes of comparison between measures of gain variously computed, Subject #48 was dropped throughout the analysis.

From Table IX, it may be seen that corresponding measures of gain for the three methods of computing gain correlated quite substantially. For example, Absolute Gain 1 correlated .96 with Proportion of Possible Gain 1, and .87 with Residual Gain 1.
### TABLE IX
INTERCORRELATIONS BETWEEN MEASURES OF GAIN (N = 76)

<table>
<thead>
<tr>
<th>Absolute Gain</th>
<th>Proportional Gain</th>
<th>Residual Gain</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(1)</td>
<td>(2)</td>
</tr>
<tr>
<td>Absolute Gain</td>
<td></td>
<td></td>
</tr>
<tr>
<td>(1)</td>
<td>1.00*</td>
<td>-.22</td>
</tr>
<tr>
<td>(2)</td>
<td>1.00*</td>
<td>-.20</td>
</tr>
<tr>
<td>Proportional Gain</td>
<td></td>
<td></td>
</tr>
<tr>
<td>(1)</td>
<td>1.00*</td>
<td>-.22</td>
</tr>
<tr>
<td>(2)</td>
<td>1.00*</td>
<td>-.21</td>
</tr>
<tr>
<td>Residual Gain</td>
<td></td>
<td></td>
</tr>
<tr>
<td>(1)</td>
<td>1.00*</td>
<td>-.04</td>
</tr>
<tr>
<td>(2)</td>
<td>1.00*</td>
<td></td>
</tr>
</tbody>
</table>

* P < .05.
It is of special interest that none of the coefficients of correlation between Gain 1 and Gain 2, for the three methods of computing gain, were positive and significant. Both positive and substantial correlation coefficients between Gain 1 and Gain 2 are a necessary condition for the demonstration of the reliability of gain measures employed. Almost the reverse of this condition was in fact the case. All of the coefficients were negative, though none within any one measure of computing gain were statistically significant at the .05 level.

**Coefficients of Correlation for Weighted Absolute Gain**

The correlation coefficients between weighted absolute gain scores and ability measures are presented in Table X. A persistent pattern prevails throughout. Whenever the measure of absolute gain was calculated with the score on the "Pretest" as the initial score (i.e., Gain 1, Gain 3, Gain 4), the coefficients between the measure of gain and ability test scores was positive and significant at the .05 level. For example, Gain 1 ("Learning Test 1" - "Pretest") correlated .58 with the DAT-VR, .49 with the CTMM-L, .56 with the DAT-VR/NA, and .38 with the CTMM-TOT. However, whenever the measure of gain was calculated with "Learning Test 1" as the initial score, the coefficient of correlation did not reach statistical significance at .05 level. For example, Gain 2 correlated .13 with
<table>
<thead>
<tr>
<th></th>
<th>Gain 1</th>
<th>Gain 2</th>
<th>Gain 1/2</th>
<th>Gain 4</th>
<th>Gain 5</th>
<th>DAT-VR</th>
<th>CTMM-L</th>
<th>DAT-VR/NA</th>
<th>CTMM-TOT</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gain 1(1)</td>
<td>1.00*</td>
<td>-.27*</td>
<td>.71* 1</td>
<td>.67*</td>
<td>-.28*</td>
<td>.58*</td>
<td>.49*</td>
<td>.56*</td>
<td>.38*</td>
</tr>
<tr>
<td>Gain 2(2)</td>
<td>1.00*</td>
<td>.49*</td>
<td>.34*</td>
<td>.75*</td>
<td>.13</td>
<td>.15</td>
<td>.14</td>
<td>.05</td>
<td></td>
</tr>
<tr>
<td>Gain 1/2(3)</td>
<td>1.00*</td>
<td>.86*</td>
<td>.30*</td>
<td>.62*</td>
<td>.62*</td>
<td>.60*</td>
<td>.35*</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Gain 4(4)</td>
<td>1.00*</td>
<td>.52*</td>
<td>.50*</td>
<td>.48*</td>
<td>.46*</td>
<td>.31*</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Gain 5(5)</td>
<td>1.00*</td>
<td>-.23</td>
<td>.06</td>
<td>.04</td>
<td>-.03</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>DAT-VR(6)</td>
<td></td>
<td>1.00*</td>
<td>.75*</td>
<td>.92*</td>
<td>.60*</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CTMM-L(7)</td>
<td></td>
<td></td>
<td>1.00*</td>
<td>.73*</td>
<td>.78*</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>DAT-VR/NA(8)</td>
<td></td>
<td></td>
<td></td>
<td>1.00*</td>
<td>.60*</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CTMM-TOT (9)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1.00*</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

* P < .05.

1 Underlined coefficients of correlation measure part-whole relations and, thus, are probably spuriously high.
Gain 4 was calculated as a measure of the amount of gain maintained over a 24-hour period (i.e., "Retention Test"—"Pretest"). In all cases the correlation between Gain 4 and the measures of ability was less than the correlation between Gain 1/2 (immediate total gain) and ability measures.

**Coefficients of Correlation for Proportion of Possible Gain**

Subject number 48 was dropped from the calculation of proportion of Possible Gain scores and from the subsequent analysis because computation of the measures of gain were not possible for this subject as previously noted.

The correlation coefficients between the measures of Proportion of Possible Gain, and between those measures and ability test scores are presented in Table XI. As may be seen, the general pattern of the coefficients corresponds to that found for Absolute Gain. Whenever Proportion of Possible Gain was computed with the "Pretest" score taken as the initial score, the relationship with ability measures was positive, significant and moderate to high. For example, Gain 1 correlated .52 with the DAT-VR, .47 with the CTMM-L, .56 with the DAT-VR/NA, and .38 with the CTMM-TOT. On the other hand, whenever Proportion of Possible Gain was computed with the scores on "Learning Test I" as the initial scores, the relation between the various measures of gain...
### TABLE XI

9 x 9 INTERCORRELATION MATRIX FOR LOG TRANSFORMED PROPORTION OF POSSIBLE GAIN (N = 76)

<table>
<thead>
<tr>
<th></th>
<th>Gain 1 (1)</th>
<th>Gain 2 (2)</th>
<th>Gain 1/2 (3)</th>
<th>Gain 4 (4)</th>
<th>Gain 5 (5)</th>
<th>DAT-VR (6)</th>
<th>CTMM-L (7)</th>
<th>DAT-VR/NA (8)</th>
<th>CTMM-TOT (9)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gain 1 (1)</td>
<td>1.00*</td>
<td>-.22</td>
<td>.69*(^3)</td>
<td>.66*</td>
<td>-.23</td>
<td>.52*</td>
<td>.47*</td>
<td>.56*</td>
<td>.38*</td>
</tr>
<tr>
<td>Gain 2 (2)</td>
<td>1.00*</td>
<td>.44*</td>
<td>.30*</td>
<td>.35*</td>
<td>.13</td>
<td>.14</td>
<td>.12</td>
<td>.05</td>
<td></td>
</tr>
<tr>
<td>Gain 1/2 (3)</td>
<td>1.00*</td>
<td>.81*</td>
<td>.26*</td>
<td>.59*</td>
<td>.51*</td>
<td>.58*</td>
<td>.32*</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Gain 4 (4)</td>
<td>1.00*</td>
<td>.41*</td>
<td>.44*</td>
<td>.40*</td>
<td>.41*</td>
<td>.25*</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Gain 5 (5)</td>
<td>1.00*</td>
<td>.03</td>
<td>.04</td>
<td>.05</td>
<td>.03</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>DAT-VR (6)</td>
<td>1.00*</td>
<td>.75*</td>
<td>.91*</td>
<td>.61*</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CTMM-L (7)</td>
<td>1.00*</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>.78*</td>
<td></td>
</tr>
<tr>
<td>DAT-VR/NA (8)</td>
<td>1.00*</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CTMM-TOT (9)</td>
<td>1.00*</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

1. The proportions were transformed to logarithmic values to the base 10 in order to normalize the distribution.
2. Subject 48 was dropped from the analysis (see text).
3. Part whole correlation coefficients are underlined.

*P < .05.
did not differ significantly from .00 ($P > .05$). For example, Gain 2 correlated .13 with the DAT-VR, .14 with the CTMM-L, .12 with DAT-VR/NA, and .05 with CTMM-TOT.

Again Gain 4 (maintenance of total immediate gain over a 24-hour period) correlated with ability measures consistently lower than did either Gain 3 or Gain 1 (immediate measures of gain with the "Pretest" taken as the initial score).

**Coefficients of Correlation for Residual Gain**

As pointed out earlier, residual gain is the most sensitive of the three measures of gain which were employed. Table XII presents the coefficients of correlation for the residual gain scores. It is apparent from Table XIII that the correlation coefficient between Gain 1 (gain with first presentation of "comprehension material") and Gain 2 (gain with second presentation of the "comprehension material") did not differ significantly from .00 ($P > .05$). A positive and significant correlation between these two measures of gain is a necessary condition for demonstrating reliability of gains. However, the correlation between Residual Gain 1 (gain with first presentation of the tape) and Residual Gain 1/2 (gain with both presentations of the tape) was .74 when computed by the formula for determining the correlation of parts with wholes (see Guilford, 1965, p. 351). Using the same formula, the correlation between Residual Gain 2 (gain with the second presentation of the tape) and Residual Gain 1/2 (gain with both
TABLE XII

9 X 9 INTERCORRELATION MATRIX FOR RESIDUAL GAIN (N = 77)

<table>
<thead>
<tr>
<th></th>
<th>Gain 1 (1)</th>
<th>Gain 2 (2)</th>
<th>Gain 1/2 (3)</th>
<th>Gain 4 (4)</th>
<th>Gain 5 (5)</th>
<th>DAT-VR (6)</th>
<th>CTMM-L (7)</th>
<th>DAT-VR/NA (8)</th>
<th>CTMM-TOT (9)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gain 1</td>
<td>1.00</td>
<td>-0.01</td>
<td><strong>0.66</strong></td>
<td>0.61</td>
<td>-0.01</td>
<td>0.61</td>
<td>0.48</td>
<td>0.61</td>
<td>0.42</td>
</tr>
<tr>
<td>Gain 2</td>
<td>1.00*</td>
<td>0.72*</td>
<td>0.57*</td>
<td>0.74*</td>
<td>0.27*</td>
<td>0.27*</td>
<td>0.28*</td>
<td>0.10</td>
<td></td>
</tr>
<tr>
<td>Gain 1/2</td>
<td>1.00*</td>
<td>0.84*</td>
<td><strong>0.54</strong></td>
<td>0.63*</td>
<td>0.54*</td>
<td>0.70*</td>
<td>0.36*</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Gain 4</td>
<td>1.00</td>
<td><strong>0.79</strong></td>
<td>0.50*</td>
<td>0.46*</td>
<td>0.49*</td>
<td>0.33*</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Gain 5</td>
<td>1.00*</td>
<td>0.15</td>
<td>0.20</td>
<td>0.16</td>
<td>0.09</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>DAT-VR</td>
<td>1.00*</td>
<td>0.75*</td>
<td><strong>0.91</strong></td>
<td><strong>0.61</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CTMM-L</td>
<td>1.00*</td>
<td><strong>0.73</strong></td>
<td>0.78*</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>DAT-VR/NA</td>
<td>1.00*</td>
<td>0.60*</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CTMM-TOT</td>
<td>1.00*</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

1Underlined coefficients of correlation indicate part-whole relationships.

*P < .05.
presentations of the tape) was found to be .66. Thus, the first and second measure of residual gain did not correlate significantly with each other, whereas, both measures of residual gain correlated appreciably with total immediate gain (Residual Gain 1/2) of which each was a part.

As with the corresponding measures of absolute gain and proportion of possible gain, Residual Gain 1 and Residual Gain 1/2 correlated positively and generally appreciably with the measures of ability. For example, Residual Gain 1 correlated .61 with the DAT-VR, .48 with the CTMM-L, .61 with the DAT-VR/NA scores, and .42 with the CTMM-TOT. When residual gain was the measure of gain employed, correlation coefficients between Gain 1 and ability measures were generally higher than when Gain 1 was computed as absolute or proportion of possible gain. As previously mentioned, this is to be expected in that residual gain is the most sensitive of the measures of gain.

Table XIII also reveals that the correlation coefficients between Residual Gain 2 and the measures of ability were statistically significant (P<.05) with the exception of the correlation of Residual Gain 2 with the CTMM-TOT. Residual Gain 2 correlated .27 with the DAT-VR, .27 with the CTMM-L, and .28 with the DAT-VR/NA scores. However, the correlation between Residual Gain 2 and the CTMM-TOT was .10 (P > .05).
As might be expected, when Gain 1 and Gain 2 were combined in an unweighted fashion (Gain 1/2, or total immediate residual gain), the correlations between this latter measure of gain and measures of ability, with the exception of the correlation with the CTMM-TOT, were higher than the correlations of either components of this measure of gain with the ability scores. Gain 1/2 correlated .61 with the DAT-VR, .48 with the CTMM-L, .61 with the DAT-VRfNA, and .42 with the CTMM-TOT.

Moreover, when Residual Gain 1 and Residual Gain 2 were combined in multiple regression analysis with ability measures as criterion variables (see Table XIII), the multiple correlation coefficients were: .67 with the DAT-VR; .55 with the CTMM-L; .68 with the DAT-VRfNA; and .43 with the CTMM-TOT.

As was the case with previous methods of computing gains, the coefficients of correlation between Residual Gain 4 (a measure of maintenance of gain over a 24-hour period) and measures of ability were consistently lower than the relation between either corresponding measure of immediate gain (Gain 1 and Gain 1/2) and ability measures. Likewise, Residual Gain 5, which measured the maintenance of Gain 2 over a 24-hour period, correlated consistently smaller with ability measures than did Gain 2.
TABLE XIII
MULTIPLE CORRELATION COEFFICIENTS BETWEEN RESIDUAL GAIN 1 AND RESIDUAL GAIN 2 CONSIDERED AS PREDICTORS, AND MEASURES OF ABILITY AS THE CRITERION

<table>
<thead>
<tr>
<th>Ability Score (Y)</th>
<th>Multiple Correlation Coefficients</th>
<th>Standard Error of R</th>
<th>Standard Error of the Estimate</th>
</tr>
</thead>
<tbody>
<tr>
<td>DAT-VR</td>
<td>.67</td>
<td>.06</td>
<td>6.86</td>
</tr>
<tr>
<td>CTMM-L</td>
<td>.55</td>
<td>.08</td>
<td>13.04</td>
</tr>
<tr>
<td>DAT-VR/ANA</td>
<td>.68</td>
<td>.06</td>
<td>13.04</td>
</tr>
<tr>
<td>CTMM-TOT</td>
<td>.43</td>
<td>.09</td>
<td>11.44</td>
</tr>
</tbody>
</table>
CHAPTER IV

THE DISCUSSION

One of the goals established for the present study was that the sample employed be as heterogeneous as possible relative to measured ability. Evidence for the achievement of this goal may be derived from Table II. It may be seen that the standard deviation of ability test scores of the present sample rather closely corresponded to the standard deviation of the national samples as reported in the California Test of Mental Maturity and Differential Aptitude Tests.

A second goal was that the learning task employed be sufficiently complex to avoid ceiling effects often present in simple tasks. Evidence that this goal was achieved may be derived from the fact that only one subject achieved a perfect score on any of the "Listening Comprehension Tests." Further, the "Listening Comprehension Tests" and the "comprehension material" were based on materials from the Advanced Test of the Iowa Silent
Reading Test. This test is appropriate for grades nine through college. Subjects in the sample employed were in the last month of the eighth grade. Therefore, it does not seem likely that the learning task suffered from ceiling effects due to simplicity of the task. Another type of ceiling effect may have operated, however. The analysis of the differences between mean scores on the "Listening Comprehension Tests" indicated that the mean score on "Learning Test II" did not differ significantly from the mean score on "Learning Test I." Thus, the average Gain 2 did not differ significantly from zero. Considering only the mean Gain 2, one is perhaps led to conclude that subjects reached a practical ceiling perhaps in the form of a plateau, or that they gained about as much as they were going to with the first presentation of the "comprehension material" and that the second presentation of the "comprehension material" had little or no effect upon learning.

However, considering the size of the standard deviation, a second interpretation seems tenable. Part of the group of subjects may have continued to gain with the second presentation of the "comprehension material," whereas, the remainder of the subjects either failed to gain or lost in performance on "Learning Test II." Such a situation might result if "Learning Test II" were more difficult than "Learning Test I." In such a case, small actual gains in learning would be counteracted by increased difficulty of
the test and computed gain scores for these subjects would be negative. Two sources of evidence cast doubt upon this second interpretation. The mean scores on "Learning Test I" and "Learning Test II" for the Item Pretest sample were equivalent. Also, Residual Gain 1 and Residual Gain 2 correlated -.01 with each other though both correlated significantly with ability measures.

A third interpretation is cautiously presented as the most tenable. It is suggested that the length of the experimental session and factors of attention and fatigue contributed to the variance of Gain 2. Thus, while some "persistent" subjects continued to gain in performance on "Learning Test II," others may have demonstrated no gains or "negative" gains in performance due to factors not directly related to learning. This interpretation is given credence by the psychological "uniqueness" of Gain 1 and Gain 2.

The above interpretation is admittedly more speculation than it is a conclusion based directly on the results of the present study. However, several writers (French, 1958; Hall, 1936; Hayes, 1962) have suggested that personality attributes such as motivational factors, should be included in the realm of that which we term intelligent behavior. It would be of interest to investigate the relation between Gain 1 and Gain 2, and measures of attributes such as motivation and persistence.
Before considering implications of the relation between gain scores and measures of ability, it is perhaps advisable to consider a possible limitation bearing upon any conclusions which may be drawn from the present study. That limitation centers around the correlation coefficients between the "Pretest" and the other "Listening Comprehension Tests," and between the "Pretest" and the measures of ability. The "Pretest" was originally designed to measure the amount of information directly relating to the "listening comprehension material" which the subject brought with him to the experimental situation. In other words, the purpose of the "Pretest" was to establish a base measure for calculating certain measures of gain. It would seem reasonable that subjects should differ in their general knowledge concerning the processing and manufacture of glass, rubber, cork and iron. We might, therefore, expect scores on a test designed to measure such general knowledge to correlate significantly with measures of ability and with tests administered after comprehension materials were presented. We might also expect the average score on such a pretest to be significantly greater than zero. Neither of these conditions held true in the present study.

On the other hand, the information presented in the "comprehension material" and the questions presented in the "Listening Comprehension Tests" were specific in nature, and
were based on the **Advanced Form of the Iowa Silent Reading Test**. Although, we may expect at least some individuals to possess previous general knowledge relevant to the topics covered, few, if any, eighth-graders would be expected to have the necessary specific information at hand in order to answer the specific questions presented on the "Pretest." Following this line of reasoning, we would not expect the "Pretest" to correlate significantly with either later "Listening Comprehension Tests" or measures of ability.

Of the three measures of gain employed in the present study, it was expected that residual gain would be the most sensitive. Residual gain measures that portion of final status which is statistically independent of initial status. If the relative sensitivity of the measures of gain may be judged by the magnitude of the correlation coefficients between measures of gain and ability, then it may be concluded that residual gain was the most sensitive measure employed.

It was somewhat surprising that, judged by the above criterion, proportion of possible gain was not a more sensitive measure than the "cruder" measure of absolute gain. However, proportion of possible gain is generally a sensitive measure of gain when the task employed is simple and suffers from "ceiling" effects. As previously discussed, this was probably not the case in the present study.
Throughout the remainder of the discussion, primary consideration will be given to measures of residual gain as they relate to measured abilities. It is first noteworthy that Residual Gain I correlated appreciably with ability test scores. In fact, whenever the measure of gain included this initial gain and was, thus, calculated with the "Pretest" score taken as the initial score, the correlation with ability measures was significant and generally fairly high. Whereas, whenever measures of residual gain did not include the "Pretest" score as the initial score, the correlation between measured gain was either non-significant or so slight that only a limited portion of the total variance of the ability measure could be accounted for by the gain measure.

On the one hand, it could be argued that the "Pretest" was not effective in measuring the amount of pertinent previous information which subjects had and that Residual Gain I was not a true or pure measure of gain, but reflected previous experience and was, thus, partially a measure of achievement. As previously discussed, the specific nature of the "comprehension material" and the questions on the "Listening Comprehension Tests" would make the extent to which the test might reflect previous experience at least questionable. Also, it is doubtful that any of the subjects had previous exposure to the Advanced Form of the Iowa Silent Reading Test since the subjects had not yet entered
the ninth grade and the test is appropriate for grades nine through college.

On the other hand, it may be that individuals learn most with the first presentation of comprehension type material. The material is new and perhaps has a certain interest appeal. With subsequent presentations of the same material, little new learning may generally occur. A hypothesis of "usefulness" of information may be forwarded. With any comprehension material, some information contained in the material may be said to be "useful," whereas, other information may be judged to be trivial.

In tests of comprehension, at least some questions are probably rather trivial in nature. Thus, a selective learning process may occur. It may be that most individuals attend to and retain "useful" information, and less "useful" information is learned and retained incidentally or not at all. Following this line of speculation, it is conceivable that the majority of useful information is learned with the first presentation of comprehension materials and subsequent presentations result in little gain in performance on the comprehension tests since little "useful" information is left.

Another of the original goals of the present study was to measure the relation between relatively permanent gain and measures of ability. It was reasoned that although
more "intelligent" persons might not be superior in making immediate gains, they might assimilate new information into their total repertoire of knowledge more efficiently and retain gains made more effectively than would less "intelligent" persons. In the present study, relative permanent gains were measured by Gain 4 and Gain 5. However, the magnitude of the correlations between these measures of gain and ability was generally smaller than was the magnitude of the correlations between corresponding measures of immediate gains and ability. Based on these results, it could be concluded that, in the present study, measures of relatively permanent gains were not more effective than measures of immediate gains in discriminating between levels of measured intellectual ability. The generality of this conclusion is limited by the fact that the "Retention Test" was given 24 hours after the last "Listening Comprehension Test" and, due to practical limitations of the item pool, the "Retention Test" was actually the last "Listening Comprehension Test" repeated. The relatively short duration of the interval between administrations of the test makes the extent to which even relatively permanent gain was measured questionable.

As pointed out in the introductory section, a persistent problem in investigations involving measures of gains and their relation to measured ability has been the unreliability of gain scores. Several investigators (Hall, 1936, Tilton, 1949 and 1953) have sought to introduce experimental
controls designed to increase the reliability of measures of gain. The apparent assumption behind such attempts is that the underlying variable, ongoing learning, is a general unitary factor, and as such is a reliable variable; whereas, measures of this variable, gain scores, are unreliable. In the study presently reported, unreliability of measures of gain was apparent. Residual Gain 1 correlated -.01 with Residual Gain 2, although both scores correlated appreciably with the composite measure of immediate residual gain of which each was a component.

An explanation of the unreliability of gain scores may be found by considering the nature of the measuring processes. Cronbach (1960) defines a test as a "sample of behavior." When performance on achievement type tests is the performance variable being measured, the measures are generally reliable. However, when the variable measured is gain in performance, unreliability of measurement is generally the case. It is suggested that achievement is a relatively stable variable which is analogous to physical attributes such as height and weight. When the variable measured is a stable one, a limited amount of performance when sampled may give a nearly "true" measure. Consider now the proposed analogous relation to physical attributes. If the relative stature and weight of a group of fourteen-year-olds were measured repeatedly over a period of a few weeks or a few months, one would
expect the measures to be fairly reliable. If these measures were correlated with the height or weight of parents, the relation would probably be positive and significant.

Now suppose that instead of measuring height or weight, we measured gains in height or weight repeatedly over a period of several months. We know that gains in weight and height of adolescents is generally not a steady linear process. It most often occurs in spurts or, to be more precise, physical growth curves show periods of rapid acceleration and periods of leveling off. Over a short period of time, these changes are generally not constant across individuals. Therefore, we would not expect measures of gain based upon such limited amounts of the attribute sampled to correlate highly with each other (be reliable) or to correlate with an external criterion such as stature and weight of parents. However, if we measured gain in weight or height over a period of a year or two, gains would probably be reliable and they might well correlate with height and weight of parents.

Relating physical attributes and gains to psychological measures of performance and gain in performance, it may be seen that measures of achievement rather resemble measures of height and weight. Both measures reflect the cumulative effects of gains over a long period of time.

Likewise, it is suggested that physical growth and "psychological" growth are analogous. Gain in weight may
reflect hereditary, emotional, and nutrition factors. Results of studies by Fleishman and Hempel (1955) and Fleishman (1960) indicate that gain in performance with practice of a task is related to different abilities at different stages of practice. Results of the present study indicate that Residual Gain 1 and Residual Gain 2 were statistically independent while both related to total gain and to measured abilities. Both in the case of physical growth and gain in performance, limitations in the amount of the processes sampled probably result in unreliability of measurement over short periods of time; the processes lack stability and, thus, lack reliability. Even more important, gains may be unreliable when different stages of gain are correlated, precisely because different factors may be contributing most to the variance of the growth factors at these stages. Moreover, the sum total of the factors acting at different stages may provide the "truest" measure of overall growth and correlate maximally with external criteria. In the present study, total immediate gain (Residual Gain 1/2) correlated consistently higher with measures of ability than did either Residual Gain 1 or Residual Gain 2; and the correlations were again consistently raised when Residual Gain 1 and Residual Gain 2 were combined in weighted linear prediction equations. This point shall be returned to later. It has been suggested here that previous experiments finding negligible
correlations between gain with practice of a task and measured intelligence have frequently suffered from limitations in the amount of gain sampled. When this was not the case, however, appreciable relationships have generally been found. For example, Tilton (1949) measured gains in performance in achievement tests of social studies over a full year period. He found that gains correlated in the area of .50 with intelligence test scores. Tilton (1953) later correlated measures of school learning over a full year period. Grade school pupils were the Ss. He found considerable evidence for a general learning factor.

Duncanson (1967) administered a number of learning tasks and ability measures to a group of sixth-grade students over a period of eight sessions on four consecutive days. Improvement in performance on the learning task was found to be related to measures of ability.

In another study (Harootunian, 1966), the time span was limited, but gains were obtained from performance on eight learning tasks. Although the correlations between the individual tasks and intelligence test scores (CTMM and Otis B) were small, a composite of learning task scores correlated as highly with IQ test scores (.71 and .73) as the intelligence tests correlated with each other.

Insofar as larger samples of learning behavior may be desirable, cumulative records from programmed learning, as discussed by Sorenson (1963) and outlined in the introductory section, might provide most useful measures of the ongoing learning processes. Controlled samples of learning
behavior over far longer periods of time than is practical with conventional methods could be obtained from which the relation between ongoing learning and measures of ability could be analyzed.

Finally, the interrelations between measures of immediate residual gain and intelligence test scores shall be considered as they relate to batteries of factorized tests and multiple prediction. Guilford has pointed out that "in building a battery of tests to predict a criterion, test makers should try to maximize the validity of each test and to minimize the correlations between tests" (1965, p. 403). It is interesting that the interrelationships between Residual Gain 1, Residual Gain 2 and measures of ability resemble the type of intercorrelations one typically finds in the case of test batteries. The two measures of gain did not correlate with each other (-.01) while both measures correlated appreciably with an unweighted composite of the scores (.74 for Residual Gain 1 and .66 for Residual Gain 2). Both contributed appreciably to the total variance of the unweighted composite of the measures (56% for Residual Gain 1 and 44% for Residual Gain 2). Furthermore, both gain measures correlated significantly with measures of verbal ability, although the correlation between Gain 1 and ability measures was sizeable (e.g., .61 with DAT-VR) and the correlations between Gain 2 and measured verbal ability was rather slight (e.g., .27
with DAT-VR and CTMK-L.

An unweighted combination of the two measures of gain correlated .61 with the DAT-VR and .48 with the CTMK-L. When combined in a regression equation, the two measures of gain correlated .68 with DAT-VR and .55 with CTMK-L.

Thus, the measures of immediate gain apparently reflected different types of improvement as indicated by the extent to which the gain measures were unique. This finding is in agreement with the results of studies by Fleishman (1960) and Fleishman and Hampel (1955) that different factors account for gains made at different stages of practice. The finding is likewise relevant to the problem of unreliability of gain scores. If gains made at different stages of practice are independent, then we should perhaps speak of gains (plural) with practice and consider gain as a composite of a number of different types of gain factors. When a measure of composite gain is desired, gain should be sampled over a long period of time or in a number of situations.

In the present study, both measures of immediate residual gain related significantly to measured verbal ability, as did a weighted and unweighted composite gain. To some extent, this finding may re-affirm our confidence in our measures of ability.
CHAPTER V

SUMMARY AND CONCLUSIONS

The present study was undertaken to investigate the relation between measured intelligence and the ability to learn as measured by gains in performance on comprehension type tests with repeated exposure to verbal comprehension material. Previous research in the area has been divided concerning the relationship between intelligence and learning ability. Certain methodological difficulties were noted in several previous investigations. Frequently the sample employed suffered from severe restrictions in the range of talent. Tasks used have generally been extremely simple, and as such may little resemble the types of material presented on intelligence tests. Also, simple tasks may suffer from "ceiling effects" which result when subjects who begin closer to the upper limit of performance have less remaining possible gain. Generally, reliabilities of gain scores have either not been reported or they were extremely low. Several measures of gain have been employed.

The present study sought to correct these methodological difficulties. A total of 77 public school eighth
grade students served as subjects in the experiment. The task employed may be described as a complex verbal comprehension task. Gains in performance were measured by alternate forms of a verbal comprehension test. The comprehension material and alternate forms of the test were presented via audio-tape in order to assure that all subjects were exposed to the complete comprehension material and had an opportunity to attempt all questions on the alternate forms of the comprehension tests and, thus, to assure that the comprehension tests were power tests.

Subjects were initially presented with a pretest in order to measure the amount of information relevant to the comprehension material which they might possess prior to the actual presentation of the material. In all, the comprehension material was presented twice via a tape recorder. After each presentation, an alternate form of the comprehension material was administered. The initial experimental session was then complete. After an interval of 24 hours, the last form of the comprehension test was again presented in order to measure amount of gain retained after an interval of delay.

Three measures of gain were computed for purposes of comparison. The measures of gain were: absolute gain, which is the crude difference between scores on two tests; proportion of possible gain, which is the ratio of absolute gain to possible gain; and residual gain, which is
defined as the difference between actual final status and final status predicted from initial status.

Reliability of gains were computed by correlating gain which corresponded to the first presentation of the comprehension material with gain which coincided with the second presentation of the tape.

Results of the correlational analyses revealed that residual gain was the most sensitive measure of gain. Since residual gain is the most refined measure of gain, discussion dealt primarily with results of this measure of gain.

Results indicated that the pretest did not correlate with any of the other measures employed. It was tentatively suggested that this could be accounted for by the specific nature of the comprehension material and of the comprehension questions.

The amount of gain retained over a one-day period of delay correlated with measures of verbal ability (Verbal Reasoning from the DAT and Language IQ from the CTMM) consistently lower than did corresponding measures of immediate gain. Interpretation was limited by the short duration of delay and by the fact that, due to practical limitations in the size of the item pool, the test presented after the delay was actually a repetition of the last test from the day before.

Results of the correlational analysis for residual gain revealed that the residual gain which coincided with
the first presentation of the comprehension material correlated appreciably with measures of verbal ability (.61 with the DAT Verbal Reasoning and .48 with the CTMM Language IQ). Residual gain which coincided with the second presentation of the comprehension material correlated significantly, but only slightly with measures of verbal ability (.27 with both measures). The intercorrelation between the two gain measures was -.01 indicating that different kinds of gain were being measured. However, both measures of gain correlated appreciably (.74 and .66) with the residual gain which coincided with both presentations of the comprehension material, a composite measure of gain of which the two measures of residual gain were the components. When gains with the first and gain with the second presentations of the comprehension material were combined in a multiple regression equation, the multiple correlation with the Verbal Reasoning subtest of the DAT was .67, while the multiple correlation with the Language IQ of the CTMM was .55.

The similarity was noted between the above pattern of intercorrelations and those found on a test consisting of a battery of factorized subtests. The composite immediate gain score could be considered as a total test score, and the component gain scores could be considered as subtests in that the gain scores did not correlate with each other, while both correlated appreciably with the total measure of gain, and both correlated significantly with the
external criteria of measures of verbal ability.

Reliability of gains defined as the correlation between gain coinciding with the first presentation of the comprehension material and gain coinciding with the second presentation of the tape was not demonstrated in the present study. However, reliability so defined is a relevant issue only if gain in performance, reflecting the ongoing learning process, is considered a general unitary factor. In the present study, the intercorrelations between measures of residual gain and verbal ability indicate the presence of independent gain factors. Residual gain with the first presentation of the verbal material, and gain with second presentation of the comprehension material may each be interpreted as defining unique factors, while both types of gain correlated significantly with measures of verbal ability.

In general, from the results obtained in the present study, it is concluded that when limited measures of gain are obtained, we might best speak of gains (plural) with practice rather than gain (singular) with practice. Gain (singular) may, however, be a suitable term for a composite measure consisting of a number of gain components. When measures of immediate gains were computed using a sensitive measure of gain, a significant relation was found between these gains, which were themselves unique, and ability tests.
The relation held, both, when the gains were considered independently and in unweighted or weighted combination. Thus, it may be seen that ability measures predicted gain at each level of practice significantly above chance, whereas gain at one level of practice did predict the gain at another level of practice. Insofar as this was the case, our confidence in our measures of ability may be enhanced.
REFERENCES


Garrett, H. E. The relation of tests of memory and learning to each other and to general intelligence in a highly selected adult group. Journal of Educational Psychology, 1928, 19, 601-613.


Hull, C. L. Variability in amount of different traits possessed by the individual. *Journal of Educational Psychology*, 1927, 18, 97-104.


Johnson, O. J. A study of the relation between ability to learn and intelligence as measured by tests. *Journal of Educational Psychology*, 1923, 14, 540-544.


APPENDIX I

ITEM STATISTICS FOR THE "LISTENING COMPREHENSION TESTS" BASED UPON THE ITEM PRETEST SAMPLE
### TABLE A

**ITEM STATISTICS FOR "LISTENING COMPREHENSION TEST I" BASED ON THE ITEM PRETEST SAMPLE**

<table>
<thead>
<tr>
<th>Item Number</th>
<th>P for 8s in Pretest Sample (N = 96)</th>
<th>Point-Biserial Correlation Between Item and Pretest</th>
<th>P for 8s in Item Pretest Sample</th>
<th>Point-Biserial Correlation Between Entire Item Pretest</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>.33</td>
<td>.30</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>.30</td>
<td>.20</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>.41</td>
<td>.30</td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>.46</td>
<td>.20</td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>.50</td>
<td>.36</td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>.44</td>
<td>.30</td>
<td></td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>.42</td>
<td>.40</td>
<td></td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>.75</td>
<td>.30</td>
<td></td>
<td></td>
</tr>
<tr>
<td>9</td>
<td>.59</td>
<td>.20</td>
<td></td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>.73</td>
<td>.50</td>
<td></td>
<td></td>
</tr>
<tr>
<td>11</td>
<td>.51</td>
<td>.50</td>
<td></td>
<td></td>
</tr>
<tr>
<td>12</td>
<td>.48</td>
<td>.40</td>
<td></td>
<td></td>
</tr>
<tr>
<td>13</td>
<td>.49</td>
<td>.75</td>
<td></td>
<td></td>
</tr>
<tr>
<td>14</td>
<td>.47</td>
<td>.22</td>
<td></td>
<td></td>
</tr>
<tr>
<td>15</td>
<td>.38</td>
<td>.32</td>
<td></td>
<td></td>
</tr>
<tr>
<td>16</td>
<td>.73</td>
<td>.20</td>
<td></td>
<td></td>
</tr>
<tr>
<td>17</td>
<td>.64</td>
<td>.25</td>
<td></td>
<td></td>
</tr>
<tr>
<td>18</td>
<td>.52</td>
<td>.20</td>
<td></td>
<td></td>
</tr>
<tr>
<td>19</td>
<td>.62</td>
<td>.50</td>
<td></td>
<td></td>
</tr>
<tr>
<td>20</td>
<td>.24</td>
<td>.65</td>
<td></td>
<td></td>
</tr>
<tr>
<td>21</td>
<td>.67</td>
<td>.20</td>
<td></td>
<td></td>
</tr>
<tr>
<td>22</td>
<td>.45</td>
<td>.25</td>
<td></td>
<td></td>
</tr>
<tr>
<td>23</td>
<td>.29</td>
<td>.60</td>
<td></td>
<td></td>
</tr>
<tr>
<td>24</td>
<td>.62</td>
<td>.35</td>
<td></td>
<td></td>
</tr>
<tr>
<td>25</td>
<td>.52</td>
<td>.30</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
### TABLE B

**ITEM STATISTICS FOR "LISTENING COMPREHENSION TEST II" BASED ON THE ITEM PRETEST SAMPLE**

<table>
<thead>
<tr>
<th>Item Number</th>
<th>P for Ss in Item Pretest Sample (N = 96)</th>
<th>Point-Biserial Correlation Between Item and Entire Item Pretest</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>.60</td>
<td>.35</td>
</tr>
<tr>
<td>2</td>
<td>.62</td>
<td>.25</td>
</tr>
<tr>
<td>3</td>
<td>.50</td>
<td>.30</td>
</tr>
<tr>
<td>4</td>
<td>.58</td>
<td>.40</td>
</tr>
<tr>
<td>5</td>
<td>.42</td>
<td>.24</td>
</tr>
<tr>
<td>6</td>
<td>.36</td>
<td>.40</td>
</tr>
<tr>
<td>7</td>
<td>.52</td>
<td>.20</td>
</tr>
<tr>
<td>8</td>
<td>.36</td>
<td>.45</td>
</tr>
<tr>
<td>9</td>
<td>.74</td>
<td>.30</td>
</tr>
<tr>
<td>10</td>
<td>.56</td>
<td>.30</td>
</tr>
<tr>
<td>11</td>
<td>.50</td>
<td>.60</td>
</tr>
<tr>
<td>12</td>
<td>.42</td>
<td>.25</td>
</tr>
<tr>
<td>13</td>
<td>.84</td>
<td>.26</td>
</tr>
<tr>
<td>14</td>
<td>.32</td>
<td>.20</td>
</tr>
<tr>
<td>15</td>
<td>.31</td>
<td>.20</td>
</tr>
<tr>
<td>16</td>
<td>.29</td>
<td>.33</td>
</tr>
<tr>
<td>17</td>
<td>.75</td>
<td>.20</td>
</tr>
<tr>
<td>18</td>
<td>.77</td>
<td>.30</td>
</tr>
<tr>
<td>19</td>
<td>.64</td>
<td>.30</td>
</tr>
<tr>
<td>20</td>
<td>.49</td>
<td>.60</td>
</tr>
<tr>
<td>21</td>
<td>.38</td>
<td>.45</td>
</tr>
<tr>
<td>22</td>
<td>.66</td>
<td>.55</td>
</tr>
<tr>
<td>23</td>
<td>.59</td>
<td>.50</td>
</tr>
<tr>
<td>24</td>
<td>.34</td>
<td>.20</td>
</tr>
<tr>
<td>25</td>
<td>.62</td>
<td>.30</td>
</tr>
</tbody>
</table>
TABLE B

ITEM STATISTICS FOR "LISTENING COMPREHENSION TEST II" BASED ON THE ITEM PRETEST SAMPLE

<table>
<thead>
<tr>
<th>Item Number</th>
<th>P for Ss in Item Pretest Sample (N = 96)</th>
<th>Point-Biserial Correlation Between Item and Entire Item Pretest</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>.60</td>
<td>.35</td>
</tr>
<tr>
<td>2</td>
<td>.62</td>
<td>.25</td>
</tr>
<tr>
<td>3</td>
<td>.50</td>
<td>.30</td>
</tr>
<tr>
<td>4</td>
<td>.58</td>
<td>.40</td>
</tr>
<tr>
<td>5</td>
<td>.42</td>
<td>.24</td>
</tr>
<tr>
<td>6</td>
<td>.36</td>
<td>.40</td>
</tr>
<tr>
<td>7</td>
<td>.52</td>
<td>.20</td>
</tr>
<tr>
<td>8</td>
<td>.36</td>
<td>.45</td>
</tr>
<tr>
<td>9</td>
<td>.74</td>
<td>.30</td>
</tr>
<tr>
<td>10</td>
<td>.56</td>
<td>.30</td>
</tr>
<tr>
<td>11</td>
<td>.50</td>
<td>.60</td>
</tr>
<tr>
<td>12</td>
<td>.42</td>
<td>.25</td>
</tr>
<tr>
<td>13</td>
<td>.84</td>
<td>.26</td>
</tr>
<tr>
<td>14</td>
<td>.32</td>
<td>.20</td>
</tr>
<tr>
<td>15</td>
<td>.31</td>
<td>.20</td>
</tr>
<tr>
<td>16</td>
<td>.29</td>
<td>.33</td>
</tr>
<tr>
<td>17</td>
<td>.75</td>
<td>.20</td>
</tr>
<tr>
<td>18</td>
<td>.77</td>
<td>.30</td>
</tr>
<tr>
<td>19</td>
<td>.64</td>
<td>.30</td>
</tr>
<tr>
<td>20</td>
<td>.49</td>
<td>.60</td>
</tr>
<tr>
<td>21</td>
<td>.38</td>
<td>.45</td>
</tr>
<tr>
<td>22</td>
<td>.66</td>
<td>.55</td>
</tr>
<tr>
<td>23</td>
<td>.59</td>
<td>.50</td>
</tr>
<tr>
<td>24</td>
<td>.34</td>
<td>.20</td>
</tr>
<tr>
<td>25</td>
<td>.62</td>
<td>.30</td>
</tr>
<tr>
<td>Item Number</td>
<td>$P$ for $3a$ in Item Prettest</td>
<td>Point-Miserial Correlation Between Item and Entire Item Prettest</td>
</tr>
<tr>
<td>-------------</td>
<td>-------------------------------</td>
<td>-------------------------------------------------------------</td>
</tr>
<tr>
<td>1</td>
<td>.68</td>
<td>.30</td>
</tr>
<tr>
<td>2</td>
<td>.61</td>
<td>.30</td>
</tr>
<tr>
<td>3</td>
<td>.75</td>
<td>.50</td>
</tr>
<tr>
<td>4</td>
<td>.40</td>
<td>.35</td>
</tr>
<tr>
<td>5</td>
<td>.35</td>
<td>.25</td>
</tr>
<tr>
<td>6</td>
<td>.53</td>
<td>.30</td>
</tr>
<tr>
<td>7</td>
<td>.55</td>
<td>.25</td>
</tr>
<tr>
<td>8</td>
<td>.50</td>
<td>.30</td>
</tr>
<tr>
<td>9</td>
<td>.48</td>
<td>.20</td>
</tr>
<tr>
<td>10</td>
<td>.55</td>
<td>.20</td>
</tr>
<tr>
<td>11</td>
<td>.40</td>
<td>.30</td>
</tr>
<tr>
<td>12</td>
<td>.33</td>
<td>.45</td>
</tr>
<tr>
<td>13</td>
<td>.30</td>
<td>.20</td>
</tr>
<tr>
<td>14</td>
<td>.29</td>
<td>.22</td>
</tr>
<tr>
<td>15</td>
<td>.82</td>
<td>.50</td>
</tr>
<tr>
<td>16</td>
<td>.67</td>
<td>.20</td>
</tr>
<tr>
<td>17</td>
<td>.56</td>
<td>.23</td>
</tr>
<tr>
<td>18</td>
<td>.57</td>
<td>.26</td>
</tr>
<tr>
<td>19</td>
<td>.60</td>
<td>.40</td>
</tr>
<tr>
<td>20</td>
<td>.52</td>
<td>.25</td>
</tr>
<tr>
<td>21</td>
<td>.36</td>
<td>.55</td>
</tr>
<tr>
<td>22</td>
<td>.22</td>
<td>.48</td>
</tr>
<tr>
<td>23</td>
<td>.63</td>
<td>.32</td>
</tr>
<tr>
<td>24</td>
<td>.55</td>
<td>.45</td>
</tr>
<tr>
<td>25</td>
<td>.49</td>
<td>.48</td>
</tr>
</tbody>
</table>
APPENDIX II

PART A--TRANSCRIPTS OF "LISTENING COMPREHENSION TESTS"

PART B--"LISTENING COMPREHENSION TESTS" ANSWER SHEETS
PART A

TRANSCRIPT OF QUESTIONS FROM "LISTENING COMPREHENSION TEST I" (TEST 2)

1. The one factor in the manufacture of glass which most affects the quality of the product is?
   (a) amount of silica used;
   (b) the melting process;
   (c) the amount of lead used.

2. What usually determines the degree of hardness of the glass?
   (a) proportion of silica;
   (b) duration of heat treatment;
   (c) proportion of sodium.

3. What are tests on the substance to be used in manufacture of glass for?
   (a) heat resistance of the substance;
   (b) presence of valuable minerals;
   (c) presence of impurities.

4. What per cent of the world's supply of cork comes from Spain and Portugal?
   (a) 70 per cent;
   (b) 50 per cent;
   (c) 90 per cent.

5. How is water removed from sheet rubber?
   (a) vacuum driers are applied;
   (b) jets of hot air are applied;
   (c) the sheets are sun-baked.

6. Where is crude rubber kept before it is taken to the manufacturing plant?
   (a) in specially heated warehouses;
   (b) in a cool and dark place;
   (c) in vats of chemical preservatives.

7. How are the particles of rubber separated from liquid latex?
   (a) by allowing it to stand;
   (b) by stirring it;
   (c) by heating it.

8. How long does the cork tree usually live?
   (a) 150 years;
   (b) 50 years;
   (c) 25 years.
1. The one factor in the manufacture of glass which most affects the quality of the product is?  
   (a) amount of silica used;  
   (b) the melting process;  
   (c) the amount of lead used.

2. What usually determines the degree of hardness of the glass?  
   (a) proportion of silica;  
   (b) duration of heat treatment;  
   (c) proportion of sodium.

3. What are tests on the substance to be used in manufacture of glass for?  
   (a) heat resistance of the substance;  
   (b) presence of valuable minerals;  
   (c) presence of impurities.

4. What per cent of the world's supply of cork comes from Spain and Portugal?  
   (a) 70 per cent;  
   (b) 60 per cent;  
   (c) 90 per cent.

5. How is water removed from sheet rubber?  
   (a) vacuum driers are applied;  
   (b) jets of hot air are applied;  
   (c) the sheets are sun-baked.

6. Where is crude rubber kept before it is taken to the manufacturing plant?  
   (a) in specially heated warehouses;  
   (b) in a cool and dark place;  
   (c) in vats of chemical preservatives.

7. How are the particles of rubber separated from liquid latex?  
   (a) by allowing it to stand;  
   (b) by stirring it;  
   (c) by heating it.

8. How long does the cork tree usually live?  
   (a) 150 years;  
   (b) 50 years;  
   (c) 25 years.
9. What type of tree does cork come from?
   (a) northern spruce;
   (b) evergreen oak;
   (c) European maple.

10. What is mixed with cork to make linoleum?
    (a) tar;
    (b) plastic cement
    (c) linseed oil.

11. What is the texture of virgin cork?
    (a) smooth;
    (b) soft;
    (c) rough.

12. What is the purpose of cleaning and flattening cork strips?
    (a) improve appearance and value;
    (b) closes pores and gives better texture;
    (c) removes dirt and parasites.

13. For what is virgin cork extensively used?
    (a) bottle stoppers;
    (b) life preservers;
    (c) rustic decorations.

14. How high does the average cork tree grow?
    (a) 10 feet;
    (b) 30 feet;
    (c) 75 feet.

15. How many longitudinal incisions are made on the tree when stripping cork?
    (a) 3 or 4;
    (b) 5 or 6;
    (c) 7 or 8.

16. In the mining of iron, what is used to clear the top surface away when the iron ore is near the surface?
    (a) dynamite;
    (b) high pressure hoses;
    (c) steam shovels.

17. The iron ore deposit of the Lake Superior Region is considered:
    (a) valuable;
    (b) average;
    (c) worthless.

18. When iron ore is refined, what is the temperature of the air which is blown into the blast furnace?
    (a) two-thousand degrees;
    (b) five-thousand degrees;
    (c) ten-thousand degrees.
19. What is the process called by which large iron ore deposits are formed?
(a) ferrication;
(b) vulcanization;
(c) sedimentation.

20. What per cent of iron is found in the ore of the Lake Superior Region?
(a) fifty to sixty;
(b) sixty to seventy;
(c) seventy to eighty.

21. Where does the United States obtain most of its iron ore?
(a) Pennsylvania;
(b) West Virginia;
(c) Lake Superior Region.

22. Where is the second important iron-producing region in the United States?
(a) Pennsylvania region;
(b) Alabama region;
(c) Lake Erie Region.

23. With what do the impurities from iron ore combine?
(a) melted coke;
(b) melted quartz;
(c) melted limestone.

24. The weight of iron ore in comparison with the impurities is?
(a) heavier than the impurities;
(b) equal to the impurities;
(c) lighter than the impurities.

25. How many of the ten important iron ore deposits are of a sedimentary variety?
(a) three;
(b) seven;
(c) ten.
1. What process in the manufacture of glass formerly made it too expensive for general use?
   (a) purifying the silica;
   (b) obtaining the silica from quartz;
   (c) melting silica.

2. What effect does the quality of sand used have on the glass?
   (a) little since impurities are removed in processing;
   (b) great since it determines the quality of the glass;
   (c) important only in the production of fine crystal.

3. In the manufacture of glass, the removal of the impurity oxide of iron is accomplished by what process?
   (a) chemical treatment;
   (b) burning and washing;
   (c) melting and rapidly coding.

4. In the homes of what type of people in this country is imported glass used?
   (a) wealthy people;
   (b) foreigners;
   (c) everyone.

5. From what source is the silica obtained which is used in making Bohemian glass?
   (a) flint;
   (b) Black Sea sand;
   (c) volcanic deposits.

6. In what form is silica obtained in most modern glass factories?
   (a) volcanic rock;
   (b) sand;
   (c) quarts rock.

7. When lime is used, it tends to make glass –
   (a) harder;
   (b) stronger;
   (c) softer.

8. What is meant by the vulcanization process?
   (a) adding chemicals;
   (b) purifying rubber;
   (c) curing by heat.
9. What name is given to the elastic substance when it first comes to the factory?
   (a) dough;
   (b) crude rubber;
   (c) elastic.

10. What appearance does rubber have after being washed?
    (a) like a black gum;
    (b) like a rolled piedough;
    (c) like a piece of sheet sponge.

11. What is the nature of the milky liquid from which rubber is derived?
    (a) a true sap;
    (b) a distilled by-product;
    (c) a secretion.

12. What is the name of the process by which rubber articles are made tough and hard?
    (a) pressurization;
    (b) solidification;
    (c) vulcanization.

13. Where is sulphur added to the raw rubber?
    (a) mixing room
    (b) chemical treatment department;
    (c) finishing room.

14. Manufactured rubber articles are first fashioned in the -
    (a) mixing room;
    (b) rubber mill;
    (c) products factory.

15. Why are special chemicals added to the rubber in the mixing room?
    (a) to dry it;
    (b) to purify it;
    (c) to vary the quality.

16. What is the first step in the treatment of the raw rubber at the factory?
    (a) it is cleaned and tested;
    (b) sulphur and other ingredients are added;
    (c) it is heat treated.

17. What is the first yield of cork called?
    (a) initial cork;
    (b) virgin cork;
    (c) black cork.
18. Cork is used extensively in the manufacture of which of the following products?
   (a) linoleum;
   (b) asphalt;
   (c) vinyl.

19. How high on the tree is cork removed?
   (a) 10 feet;
   (b) just below main branches;
   (c) half-way up the tree.

20. What effect does stripping the cork have on the trees?
   (a) detrimental effect;
   (b) no effect;
   (c) beneficial effect.

21. What quality of cork makes it satisfactory for bottle stoppers?
   (a) solidness;
   (b) cheapness;
   (c) elasticity.

22. What substances are used in refining iron ore?
   (a) limestone and coke;
   (b) sand and charcoal;
   (c) quartz and lime.

23. When iron is refined, what is the temperature of the air used in the furnace?
   (a) one-thousand degrees;
   (b) five-thousand degrees;
   (c) ten-thousand degrees.

24. In iron-ore mines, how is the ore sent to the surface?
   (a) through shafts;
   (b) in elevators;
   (c) on conveyors.

25. In the process of refining iron ore, what are the impurities and melted limestone called?
   (a) concentrate;
   (b) slag;
   (c) pig iron.
1. What kinds of glass are still made from the silica in quartz rock?
   (a) milk glass
   (b) Bohemian glass;
   (c) Brazilian glass.

2. What is the relative cost of glass made from flint and quartz?
   (a) cheap;
   (b) average;
   (c) expensive.

3. The impurity in sand which most seriously affects the clearness of glass is -
   (a) vegetable matter;
   (b) clay;
   (c) oxide of iron.

4. What process is always used in the manufacture of glass?
   (a) crushing quartz;
   (b) compressing sand;
   (c) melting silica.

5. What substance used in making glass tends to make the glass softer?
   (a) lead;
   (b) lime;
   (c) iron.

6. What determines the transparency of the glass?
   (a) amount of lead;
   (b) quality of sand;
   (c) proportion of sand.

7. Sand for use in making glass must not contain more iron than -
   (a) one-half of one per cent;
   (b) one per cent;
   (c) five per cent.

8. How is iron oxide removed from the sand used for making glass?
   (a) using chemicals;
   (b) burning;
   (c) washing.
9. What type of substance is crude rubber?
   (a) firm and elastic;
   (b) soft and gummy;
   (c) hard and brittle.

10. Where do rubber trees usually grow?
    (a) Portugal and Spain;
    (b) temperate zone;
    (c) tropics.

11. What happens if the juice of the rubber plant is allowed to stand for a time?
    (a) particles of rubber rise to the surface;
    (b) particles of rubber settle to the bottom;
    (c) masses of rubber separate out.

12. What is used to cleanse the rubber at the factory?
    (a) hot oil;
    (b) fresh water;
    (c) strong acid.

13. What is done to change crude rubber into a putty like mass?
    (a) it is steam treated;
    (b) it is chemically treated;
    (c) it is rubbed and crushed.

14. What determines the hardness of the rubber product?
    (a) amount that rubber is compressed;
    (b) proportion of pure rubber in the product;
    (c) amount of heat applied in processing.

15. What countries grow the greatest amount of cork?
    (a) Italy and Greece;
    (b) Burma and Ceylon;
    (c) Portugal and Spain.

16. How long does the cork tree continue to produce cork?
    (a) 50 years or more;
    (b) 100 years or more;
    (c) 150 years or more.

17. Stripping the cork from the tree has what effect on the tree?
    (a) beneficial;
    (b) detrimental;
    (c) of no effect.

18. Where in this country are experiments in growing cork being conducted?
    (a) California;
    (b) Texas;
    (c) Florida.
19. What is the average yield of a tree at each cutting of cork?
   (a) 10 lbs.
   (b) 45 lbs.
   (c) 120 lbs.

20. At what time of the year is the cork taken from the tree?
   (a) summer;
   (b) winter;
   (c) spring.

21. How are the strips cared for after being taken from the trees?
   (a) surfaces are treated chemically and baked;
   (b) surfaces are wire brushed and sun dried;
   (c) surfaces are cleaned and flattened by pressure.

22. Cork is used for the soles of shoes because it is -
   (a) non-conductive;
   (b) durable;
   (c) inexpensive.

23. What is the substance that is placed in the blast furnace with the iron ore?
   (a) coke;
   (b) granite;
   (c) slag.

24. How does iron compare in weight with the impurities in the ore?
   (a) lighter;
   (b) approximately the same;
   (c) heavier.

25. What is used to make the molds into which the iron is poured?
   (a) limestone;
   (b) sand;
   (c) slag.
PART B
"LISTENING COMPASSION TEST" ANSWER SHEETS

NAME

SAMPLE QUESTION

1. (a) rock; (b) iron; (c) sand.
1. (a) amount of silica used; (b) the melting process; (c) the amount of lead used.

2. (a) proportion of silica; (b) duration of heat treatment; (c) proportion of sodium.

3. (a) heat resistance of the substance; (b) presence of valuable minerals; (c) presence of impurities.

4. (a) 70 per cent; (b) 80 per cent; (c) 90 per cent.

5. (a) vacuum dryers are applied; (b) jets of hot air are applied; (c) the sheets are sun baked.

6. (a) in specially heated warehouses; (b) in a cool and dark place; (c) in vats of chemical preservatives.

7. (a) by allowing it to stand; (b) by stirring it; (c) by heating it.

8. (a) 150 years; (b) 50 years; (c) 25 years.

9. (a) northern spruce; (b) evergreen oak; (c) European maple.

10. (a) tar; (b) plastic cement; (c) linseed oil.

11. (a) smooth; (b) soft; (c) rough.

12. (a) improve appearance and value; (b) closes pores and gives better texture; (c) removes dirt and parasites.

13. (a) bottle stoppers; (b) life preservers; (c) rustic decorations.

14. (a) 10 feet; (b) 30 feet; (c) 75 feet.

15. (a) 3 or 4; (b) 5 or 6; (c) 7 or 8.

16. (a) dynamite; (b) high pressure hoses; (c) steam shovels.

17. (a) valuable; (b) average; (c) worthless.

18. (a) two thousand degrees; (b) five thousand degrees; (c) ten thousand degrees

19. (a) ferrication; (b) vulcanization; (c) sedimentation.

20. (a) fifty to sixty; (b) sixty to seventh; (c) seventy to eighty.

21. (a) Pennsylvania; (b) West Virginia; (c) Lake Superior Region.

22. (a) Pennsylvania region; (b) Alabama region; (c) Lake Erie Region.

23. (a) melted coke; (b) melted quartz; (c) melted limestone.

24. (a) heavier than the impurities; (b) equal to the impurities; (c) lighter than the impurities.

25. (a) three; (b) seven; (c) ten.
1. (a) purifying the silica; (b) obtaining the silica from quartz; (c) melting silica.

2. (a) little since impurities are removed in processing; (b) great since determines the quality of the glass; (c) important only in the production of fine crystal.

3. (a) chemical treatment; (b) burning and washing; (c) melting and rapidly cooling.

4. (a) wealthy people; (b) foreigners; (c) everyone.

5. (a) flint; (b) Black Sea sand; (c) volcanic deposits.

6. (a) volcanic rock; (b) sand; (c) quartz rock

7. (a) harder; (b) stronger; (c) softer.

8. (a) adding chemicals; (b) purifying rubber; (c) curing by heat.

9. (a) dough; (b) crude rubber; (c) elastic.

10. (a) like a black gum; (b) like a rolled pie dough; (c) like a piece of sheet sponge.

11. (a) a true sap; (b) a distilled by-product; (c) a secretion.

12. (a) pressurization; (b) solidification; (c) vulcanization.

13. (a) mixing room; (b) chemical treatment department; (c) finishing room.

14. (a) mixing room; (b) rubber mill; (c) products factory.

15. (a) to dry it; (b) to purify it; (c) to vary the quality.

16. (a) it is cleaned and tested; (b) sulphur and other ingredients are added; (c) it is heat treated.

17. (a) initial cork; (b) virgin cork; (c) black cork.

18. (a) linoleum; (b) asphalt; (c) vinyl.

19. (a) 10 feet; (b) just below main branches; (c) half way up the tree.

20. (a) detrimental effect; (b) no effect; (c) beneficial effect.

21. (a) solidness; (b) cheapness; (c) elasticity.

22. (a) limestone and coke; (b) sand and charcoal; (c) quartz and lime.

23. (a) one thousand degrees; (b) five thousand degrees; (c) ten thousand degrees.

24. (a) through shafts; (b) in elevators; (c) on conveyors.

25. (a) concentrate; (b) slag; (c) pig iron.
1. (a) milk glass; (b) Bohemian glass; (c) Brazilian glass.
2. (a) cheap; (b) average; (c) expensive.
3. (a) vegetable matter; (b) clay; (c) oxide of iron.
4. (a) crushing quartz; (b) compressing sand; (c) melting silica.
5. (a) lead; (b) lime; (c) iron.
6. (a) amount of lead; (b) quality of sand; (c) proportion of sand.
7. (a) one half of one per cent; (b) one per cent; (c) five per cent.
8. (a) using chemicals; (b) burning; (c) washing.
9. (a) firm and elastic; (b) soft and gummy; (c) hard and brittle.
10. (a) Portugal and Spain; (b) temperate zone; (c) tropics.
11. (a) particles of rubber rise to the surface; (b) particles of rubber settle to the bottom; (c) masses of rubber separate out.
12. (a) hot oil; (b) fresh water; (c) strong acid.
13. (a) it is steam treated; (b) it is chemically treated; (c) it is rubbed and crushed.
14. (a) amount that rubber is compressed; (b) proportion of pure rubber in the product; (c) amount of heat applied in processing.
15. (a) Italy and Greece; (b) Burma and Ceylon; (c) Portugal and Spain.
16. (a) 50 years or more; (b) 100 years or more; (c) 150 years or more.
17. (a) beneficial; (b) detrimental; (c) of no effect.
18. (a) California; (b) Texas; (c) Florida.
19. (a) 10 lbs.; (b) 45 lbs.; (c) 120 lbs.
20. (a) summer; (b) winter; (c) spring.
21. (a) surfaces are treated chemically and baked; (b) surfaces are wire brushed and sun dried; (c) surfaces are cleaned and flattened by pressure.
22. (a) non-conductive; (b) durable; (c) inexpensive.
23. (a) coke; (b) granite; (c) slag.
24. (a) lighter; (b) approximately the same; (c) heavier.
25. (a) limestone; (b) sand; (c) slag.
APPENDIX III

PART A
DESCRIPTION OF I.B.M. 1620 PROGRAM EMPLOYED

PART B
FORTRAN PROGRAMS WRITTEN BY THE AUTHOR AND
USED IN THE DATA ANALYSIS
PART A

DESCRIPTION OF I.B.M. 1620 PROGRAM EMPLOYED

Title: Single and Multiple Linear Regression Analysis Program

Author: Anthony J. Capato
Chemical Engineering Dept.
Columbia University

Description: The program uses a least squares procedure to calculate the estimates of the partial regression coefficients.

The maximum number of independent variables is ten. The number of data points is unlimited. The program also computes the partial correlation coefficients, the multiple correlation coefficient, the standard error of the Y data, the standard error of the estimate, the significance of regression, and the standard error of the partial regression coefficients.

The program was written in Fortran with format.

(Abstracted from: Catalog of Programs for I.B.M. 1620 and 1700 Data Processing Systems; June, 1966, p. 57)
PART B

FORTRAN PROGRAMS WRITTEN BY THE AUTHOR AND
USED IN THE DATA ANALYSIS

1. Program to Calculate Weighted Test Scores from Coded Student Answers.
2. Program to Calculate Absolute Gain Score from Weighted Raw Scores.
3. Program to Calculate Proportion of Possible Gain Scores from Weighted Raw Scores.
4. Program to Log Transform Proportion of Possible Gain Scores.
5. Program to Calculate Residual Gain Scores.
6. Program to Calculate Means and Standard Deviations of Weighted Test Scores.
7. Program to Compute Single Classification Analysis of Variance for Repeated Measures.
VITA

Robert Stephen Feddicord was born in Baltimore, Maryland, on March 8, 1944. He attended Baltimore Public Schools and was graduated from Edmondson Senior High School on February 2, 1962. In 1966, he received his B.A. degree from Gettysburg College. He was elected to Psi Chi in 1965. After completing his work toward the M.A. degree at the University of Richmond, he expects to assume a position at the John F. Kennedy Institute for the Habilitation of Mentally and Physically Handicapped Children.