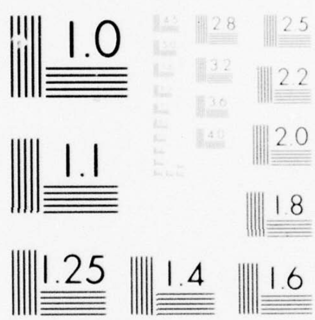
AD-A043 687 MISSOURI UNIV-KANSAS CITY EDUCATIONAL PSYCHOLOGY AND --ETC F/G 5/10
THE RELATIONSHIP BETWEEN REASONING ABILITY AND GAIN IN READING --ETC(U)
AUG 77 R P CARVER N00014-75-C-0958

UNCLASSIFIED NL

| OF |
AD
A043687

END
DATE
FILMED
9 -77
DDC



MICROCOPY RESOLUTION TEST CHART
NATIONAL BUREAU OF STANDARDS-1963-A

AD A043687

(12)

THE RELATIONSHIP BETWEEN REASONING ABILITY AND GAIN IN READING ABILITY

RONALD P. CARVER

Final Report

August 1977

DDC
RECEIVED
SEP 1 1977
B

AD No.
DDC FILE COPY

Educational Research and Psychology Division

School of Education

University of Missouri - Kansas City

DISTRIBUTION STATEMENT A
Approved for public release;
Distribution Unlimited

Unclassified

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE

READ INSTRUCTIONS BEFORE COMPLETING FORM

1. REPORT NUMBER: Final Report

2. GOVT ACCESSION NO.:

3. RECIPIENT'S CATALOG NUMBER:

4. TITLE (and Subtitle): The Relationship between Reasoning Ability and Gain in Reading Ability.

5. TYPE OF REPORT & PERIOD COVERED: Final Report, 1 Jan 1975 - 31 Aug 1977

6. PERFORMING ORG. REPORT NUMBER:

7. AUTHOR(s): Ronald P. Carver

8. CONTRACT OR GRANT NUMBER(s): N00014-75-C-0958

9. PERFORMING ORGANIZATION NAME AND ADDRESS:
School of Education
University of Missouri-Kansas City
Kansas City, Missouri 64110

10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS:
61153N; RR042-04; RR 042-04-01
NR 150-369

11. CONTROLLING OFFICE NAME AND ADDRESS:
Personnel and Training Research Programs
Office of Naval Research (Code 458)
Arlington, Virginia 22217

12. REPORT DATE: 31 Aug 1977

13. NUMBER OF PAGES: 91

14. MONITORING AGENCY NAME & ADDRESS (if different from Controlling Office): 89p.

15. SECURITY CLASS. (of this report): Unclassified

15a. DECLASSIFICATION/DOWNGRADING SCHEDULE:

16. DISTRIBUTION STATEMENT (of this Report):
Approved for public release; distribution unlimited.

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report):

18. SUPPLEMENTARY NOTES:

19. KEY WORDS (Continue on reverse side if necessary and identify by block number):
Reading, Reasoning, Intelligence, PLATO IV, computer, training, programmed prose, mastery learning, Intellectual capacity, Raven test, National Reading Standards Test, readability, measuring gain

20. ABSTRACT (Continue on reverse side if necessary and identify by block number):
(see reverse side)

DD FORM 1 JAN 73 1473 EDITION OF 1 NOV 65 IS OBSOLETE

S/N 0102-LF-014-6601

Unclassified

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

410353

ABSTRACT

The general objective of the proposed research was to determine how reasoning ability, or intellectual capacity, affects gain in reading ability. The theory underlying the present research was that poor readers consist of two types--those who read poorly because of deficits in their reasoning ability and those who read poorly because of deficits in reading practice, or experience. The effect of reading practice upon reading ability was investigated using high school students who read poorly. The reading training involved a recently developed technique, called programmed prose, which allows regular reading material to be automatically converted into training material. Each programmed prose passage was read and reread until mastery (100% accuracy) was achieved. A PLATO IV computer terminal was used to administer the programmed prose passages. Each student was given 50 to 70 hours of individualized instruction on the terminal. There were three separate studies with six high school students in each study. All students were at about grade level 5 in reading ability prior to the instruction. One half of the students in each study were selected because they purportedly had high reasoning ability, as indicated by high scores on the Raven Progressive Matrices Test; the other one half had low scores on this test. The results were consistent across all three studies. The high Raven groups, who supposedly had high reasoning ability and should benefit greatly from reading training, did not gain more than the low Raven groups. When gain in reading ability was measured using a test that was just like the task employed in the reading training, there was a large amount of gain--from grade level 5 to 8. However, when gain in reading ability was measured using other techniques, there was little or no evidence that the training, i.e., reading practice, produced gain. Also, the Raven test was administered under special research conditions, and the results suggested that the original test results, which were used to divide the students into high and low groups, may not have been valid. This research failed to find a relationship between reasoning ability and gain in reading ability but this failure could have been due to: (a) ineffective techniques used to produce gains in reading ability, and (b) ineffective techniques used to measure reasoning ability.

THE RELATIONSHIP BETWEEN REASONING ABILITY AND GAIN IN READING ABILITY

Ronald P. Carver
Principal Investigator

FINAL REPORT

Prepared under Contract to the
Personnel and Training Research Programs
Psychological Sciences Division
Office of Naval Research
Department of the Navy

Contract No. N00014-75-C-0958
NR No. 150-369/12-13-74 (458)

Approved for public release; distribution unlimited.
Reproduction in whole or in part is permitted for
any purpose of the United States Government.

University of Missouri-Kansas City
School of Education
Educational Psychology and Research Division

August 1977

ABSTRACT

The general objective of the proposed research was to determine how reasoning ability, or intellectual capacity, affects gain in reading ability. The theory underlying the present research was that poor readers consist of two types--those who read poorly because of deficits in their reasoning ability and those who read poorly because of deficits in reading practice, or experience. The effect of reading practice upon reading ability was investigated using high school students who read poorly. The reading training involved a recently developed technique, called programmed prose, which allows regular reading material to be automatically converted into training material. Each programmed prose passage was read and reread until mastery (100% accuracy) was achieved. A PLATO IV computer terminal was used to administer the programmed prose passages. Each student was given 50 to 70 hours of individualized instruction on the terminal. There were three separate studies with six high school students in each study. All students were at about grade level 5 in reading ability prior to the instruction. One half of the students in each study were selected because they purportedly had high reasoning ability, as indicated by high scores on the Raven Progressive Matrices Test; the other one half had low scores on this test. The results were consistent across all three studies. The high Raven groups, who supposedly had high reasoning ability and should benefit greatly from reading training, did not gain more than the low Raven groups. When gain in reading ability was measured using a test that was just like the task employed in the reading training, there was a large amount of gain--from grade level 5 to 8. However, when gain in reading ability was measured using other techniques, there was little or no evidence that the training, i.e., reading practice, produced gain. Also, the Raven test was administered under special research conditions, and the results suggested that the original test results, which were used to divide the students into high and low groups, may not have been valid. This research failed to find a relationship between reasoning ability and gain in reading ability but this failure could have been due to: (a) ineffective techniques used to produce gains in reading ability, and (b) ineffective techniques used to measure reasoning ability.

ACCESSION for
NTIS White Section ☑
DDC Buff Section ☐
UNANNOUNCED ☐
JUSTIFICATION
BY
DISTRIBUTION/AVAILABILITY CODES
Dist. AVAIL. and/or SPECIAL
A

PRECEDING PAGE BLANK-NOT FILMED

ACKNOWLEDGEMENTS

The data collection and data analyses were accomplished with the assistance of Fred Duffelmeyer, James Hoffman, Mark Condon, Drew Love, Ruth Cassandras, Marc Rebstock, Greg Savage, and Frances Wenner.

Without the cooperation of the administration, faculty, and students of two Kansas City, Missouri high schools, Paseo and Southwest, this research would not have been possible. Special thanks goes to Anna Marie Peterson and Barbara Lusk of Paseo, J. C. Reneau, Principal; and Jim Woods of Southwest, Thomas Kipp, Principal.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

INTRODUCTION

In 1917, Edward L. Thorndike analyzed reading, and he argued that reading was basically a reasoning process. In 1971, Robert L. Thorndike presented correlational and factor analytic data of individual differences which he interpreted as supporting this analysis of the elder Thorndike. Robert L. Thorndike went a step further and argued that if we desire better readers, the challenge is to develop ways of teaching people to think rather than concentrating on reading. He concluded that it is primarily meager intellectual processes that are limiting reading comprehension, not deficits in one or more specific and readily teachable skills.

If reading is primarily reasoning and if individual differences in the ability to reason do account for almost all of the variation in individual differences in reading ability, then is reading training a waste of time for most poor readers? The answer to this theoretical question has important practical consequences. Many high school students and adults are extremely poor readers, i.e., they are described as functionally illiterate. The question is, are most of these people poor readers because they are poor reasoners, or thinkers? If so, then reading instruction is likely to have very little effect upon their reading ability. Policy makers in training, and education in general, should not place a great deal of resources into reading training for poor readers if these theoretical ideas of Thorndike are correct.

It would seem to be of great practical benefit to anyone concerned with education to have empirical data directly relevant to the above theoretical issue. What is the relationship between reasoning ability, or intellectual capacity, and improvement in reading ability? The empirical data collected up to now has focused upon individual differences in reading ability and how they relate to individual differences in intellectual capacity. There has been little attention given to research that is directly relevant to the issue in question, i.e., how does reasoning ability, or intellectual ability, relate to within-individual gain in reading ability.

The following parts of this introduction section will present the background for an empirical study of the relationship between reasoning ability and gain in reading ability.

REASONING ABILITY

This section will briefly present more of the theoretical background underlying the investigation. For the purposes of this research, reasoning ability, intellectual capacity, and intellectual ability will be regarded as synonomous.

Palermo and Molfese (1972) found that mental age was closely related to performance measures of language development and they argued that "...a theory of language development must be embedded within the larger context of a theory of cognitive development /p. 426/." Crano, Kenny, and Campbell (1972) carried this idea one step further when they contended that their data indicate that "...nonverbal intelligence causally influences verbal IQ, an ability which, in turn, is a predictor of many of the more concrete linguistic skills..../p. 272/." Singer (1976) argued that under certain conditions IQ was likely to have a significant relationship with the rate of acquisition of a task.

The importance of conducting research in this area has been articulated by:

(a) Harber and Bryen (1976) who wrote that "...more information is needed concerning other variables that affect reading performance, such as intelligence /p. 400/, "and (b) Lohnes and Grey (1972) who wrote that "...reading experts must understand intellegence to understand reading /p. 475/."

READING PROCESS

There is little or no doubt but that the reading process does involve interactions with such constructs as thinking, reasoning, and intellectual capacity. In the recent past a large number of "models" of the reading process have been advanced (see review by Geyer, 1973), and none of these models suggest that reading is exclusively a type of rote psychomotor performance that does not require higher order intellectual processes such as reasoning. Basic intellectual competency is generally implicitly regarded as a "given" in such discussions. There has been little concern for how individual differences in reasoning ability interact with the reading process. Some individuals may be better able to comprehend the relationships among things better than other individuals and this basic intellectual skill may simply manifest itself in the reading situation, i.e., a situation which involves the cognition of relations between words.

Carver (1973) has isolated levels of the reading process as described in the reading literature (see Spache, 1963), and has related these levels to intellectual functioning. Level 1 is associated with words as units and involves both the decoding of words and the determination of their meaning as used in the particular sentence being read. Level 2 is associated with the sentences as units and involves the combination of the meanings of the individual words into the complete understanding of the sentence. Level 3 is associated with the paragraph as a unit and may involve such processes as the recognition of the implied main idea of a paragraph. Level 4 is associated with no particular unit and may involve thinking activities which are not at all associated with the literal, implied, or tangential meanings of the prose. By definition, Levels 3 and 4 are primarily reasoning processes, i.e., intellectual functioning processes that are not inherent or specific to reading and do not occur simultaneously with the reading process itself. Therefore, it is misleading to suggest that reading is primarily reasoning when reading refers to Levels 3 and 4 because Levels 3 and 4 involve reasoning by definition. Levels 1 and 2 are inherent to the reading processes, i.e., they involve an ongoing interaction between the stimuli perceived, i.e., the words, and the cognition of the intended meaning of these stimuli. Levels 1 and 2 of reading do not involve reasoning ability or intellectual capacity by definition.

In summary, the essence of reading, i.e., the process of recognizing the intended meaning of words and sentences as in Levels 1 and 2 above, is a process which would seem to depend upon a fundamental intellectual ability such as reasoning. Therefore, it would seem that differences in reasoning ability would place limits upon the reading ability of individuals.

READING PROGRESS

The preceding discussion has been restricted to the relationship between reasoning ability and the reading process for mature readers, i.e., individuals who are primarily reading to learn rather than learning to read. In most learning to read situations, the relationship becomes more complex. This is because most of the individuals who are learning to read are also maturing intellectually so that reasoning ability and reading ability are both progressing at the same time. Such a dynamic relationship is not as simple to scrutinize and to explain. An explanation of these complex relationships can be facilitated using Carroll's (1963) model of school learning. This model will be described briefly before attempting to integrate it into the reading situation.

Carroll envisioned school learning as being influenced by five primary factors. Three of these factors were internal or individual difference factors while two were external to the individual, i.e., treatment factors. The three individual difference factors were: (1) aptitude--the amount of time needed to learn the task under optimal instructional conditions, (2) ability to understand instruction, and (3) perseverence--the amount of time the individual is willing to engage in learning. The two treatment factors, or external factors, were: (4) opportunity--time allowed for learning, and (5) the quality of instruction. Carroll regards his aptitude factor as relatively resistant to change and musters support for this stance by referring to the research which has found the IQ to be relatively constant. Carroll regards his second factor, ability to understand instruction, as a combination of general and verbal intelligence, i.e., as assessable "...in relative terms by currently available measuring devices /p. 729/." Carroll's model is an excellent one because it focuses upon measurable entities which allow his theory to be empirically tested. There is empirical support for Carroll's theory (e.g., see Carver, 1970).

The five preceding factors are considered by Carroll as influencing the amount of time the learner <u>needs</u> to master the learning task. Degree of learning is "...a function of the ratio of the amount of time the learner actually spends on the learning task to the total amount he needs /p. 730/." Other things being equal, degree of learning...is a simple function of the amount of time during which the pupil engages actively in learning /p. 732/."

POOR READERS

An attempt will now be made to apply the preceding theoretical background to a specific reading situation of great practical interest. Why do some students read poorly? It is hypothesized that, as a group, poor readers contain two general types; each type reads poorly for different reasons.

The Type I individual reads poorly because of low reasoning ability. This individual may have a chronological age of 18 but have the reasoning ability of a ten year old and read at a level equivalent of that of a 10 year old. This person only reads poorly when compared to other individuals of the same chronological age. This individual may be considered as an average reader if compared to other individuals of the same reasoning ability. The Type I poor readers are the type that Thorndike (1971) was talking about when he contended that it was primarily meager intellectual skills that were contributing to poor reading and that increases in reading ability are not likely to precede increases in thinking ability. In Carroll's model, the Type I poor reader has spent the amount of time needed in reading situations. This individual's poor reading does not result from a lack of an opportunity to read, from poor instructional quality, or from lack of personal perseverence.

The Type II individual reads poorly primarily because of a lack of reading practice. This individual has not spent the time that was necessary to read at a level that would be expected from his/her intellectual capacity. There are many possible reasons why this type of individual has not spent a sufficient amount of time reading. Preschool cultural or environmental forces may have influenced the individual to focus upon physical acitivites rather than words and language. In school, the individual may have found that he was behind and that catching up was a difficult and time consuming task. Furthermore, he may not have been willing to persevere. Upon getting behind early, the school instruction also may have been inappropriate to the individual's level. Therefore, even the effort expended by the individual did not result in the same amount of learning as peer students because the quality of the instruction was not optimal. This interaction between poor reading (in relationship to peers) and poor instruction (in relationship to peers) may have severly discouraged participation in the act of reading to the point where reading was avoided if possible. Such an effect, as described above, would multiply the disparity between an individual and his peer group. It is theorized, however, that the ability of the individual could be predicted given a knowledge of the amount of time the individual had engaged in the reading act during his school years. That is, this individual might be regarded as an average reader when his/her total amount of reading experience or practice is compared to other individuals who have spent comparable amount of time reading.

Stated differently, the Type I readers are considered to be normal achievers who have a level of reading ability which is commensurate with their level of reasoning ability. The Type II readers are underachievers who are reading below their potential primarily because they have not engaged in the amount of reading practice that would be expected from their years in school. Using the parlance of verbal learning theory, the amount of the nominal stimulus (years of school instruction) for the Type II individuals is an overestimate of the amount of the effective stimulus (years of reading practice).

There is a very important hypothesis that can be derived from Carroll's theory which is directly relevant to reading. Carroll mentioned but did not speculate upon the cumulative effects of learning different tasks and how this factor interacted with time. For example, suppose Individual A in an elementary school has a higher reasoning ability than Individual B, and therefore needs less time to master the material contained in one of the basal readers used for instruction. Individual A will therefore have completed that particular level of the basal reader series and be working on a higher level while Individual B is still working at the same level. The total time spent reading for the two individuals would be equal but Individual A would have progressed to a higher level of reading ability during a fixed amount of time. From Carroll's model it could be hypothesized that: (a) equal amounts of reading practice, from a task-mastery standpoint, would result in equal gains in reading ability no matter what the individuals' reasoning ability, and (b) equal amounts of time would produce unequal amounts of gain in reading ability depending upon the individual's reasoning ability.

The above theoretical relationship concerning task-mastery and time spent learning will now be integrated into the previously discussed theoretical difference between Type I and Type II poor readers. Suppose a group of Type I readers and a group of Type II readers were given a number of passages that they must learn to read to a certain criterion level of accuracy. The task-mastery hypothesis would predict that: (a) at the point where both groups had mastered the same number of passages, their reading ability would also have increased to an equivalent level, and (b) the Type II readers would master more passages in an equal amount of time and would therefore have reached a higher level of reading ability when the amount of time engaged in reading was equal.

In summary, it is theorized that poor readers are of two types. Both types gain equal amounts in reading ability when they have mastered equal amounts of instructional tasks. However, the Type I poor readers have a higher reasoning ability and this results in their gaining more than the Type II poor readers during equal amounts of time.

MEASUREMENT METHODS

Introduction. In order to test this hypothesized relationship between reasoning ability and gain in reading ability, valid measurement techniques must exist. This section will contain a discussion of measurement techniques to be used to provide indicators of reasoning ability, reading ability, and degree of underachievement.

Reasoning Ability. The following quotation from Robert Gagné (1973) provides a summary of knowledge about reasoning ability, or intellectual capacity:

> ...there are vast differences in the capacity for learning among different individuals. By capacity I mean to refer to genetically determined intellectual potential. The exact nature of these differences in capacity remains a mystery today, despite many years of investigation into what we call intelligence /p. 2/.

Although there may be a lack of consensus about how one should measure intellectual capacity, or reasoning ability, it is possible and reasonable to investigate hypotheses about reasoning ability using psychological tests.

At this point, it is necessary that a clear distinction be made between reasoning ability and IQ. An IQ test is generally regarded as measuring intelligence, and such tests normally measure intelligence in relationship to a normative group, e.g., same age peers. Intelligence, as measured in this manner, stays relatively constant from childhood to adulthood. However, during these years of constant IQ, reasoning ability increases each year. In this sense, reasoning ability is conceptually equivalent to what was formerly measured by Mental Age, M.A., i.e., before IQ was changed to a deviation IQ.

Existing group tests of intelligence are usually contaminated by requiring an ability to read. One group test which does not require reading and yet appears to be an excellent measure of reasoning ability is the Raven Progressive Matrices Test. It consists of 60 figures with parts missing. The task is to select one of the alternatives given which correctly completes the figure. The figures involve the understanding of relationships which vary from extremely easy to extremely difficult, so intellectual capacity can be measured from a mental age of about 8 on up to maturity. In the test manual, the Raven is alternately described as providing a measure of "reasoning," "clear thinking," and "intellectual capacity."

The Raven has been considered by Charles Spearman (1946), and others (see Vernon, 1947; Vincent, 1952; Jensen, 1969) as the best of the nonverbal tests for measuring g, general intelligence. In Guilford's (1967) structure of the intellect, the Raven is considered as measuring the Cognition of Relations between figures (Figural). Reading would be considered as measuring the Cognition of Relations between words (Semantic). That is, the Raven could be considered as involving the same type of mental ability as reading--cognition of relations--except the content would be different--figures versus words. The Raven has also been used in studies of learning proficiency (e.g., see Rohwer, Ammon, Suzuki, and Levin, 1971), and a relatively recent administration of the Raven to fraternal and identical twins indicated that about 85% of the variance of the g trait, as estimated by this test, was inherited (Pezzullo, Thorsen, & Madans, 1972). Nelson and Edelstein (1963) suggested that the "...Raven Progressive Matrices Test might offer an additional method of assessing the intelligence of children with language and/or cultural handicaps /p. 46/."

The Raven is probably the best available group test to use to estimate an individual's general intellectual ability or reasoning ability, especially when the ultimate use of the test is to estimate reading potential in a manner that does not involve words.

Reading Ability. Existing standardized tests for measuring reading achievement suffer from two interrelated deficiencies. First, the tests have been designed to be maximally efficient in reflecting individual differences in reading ability, and therefore do not allow performance to be measured on an interval or ratio scale. Second, by selecting items which best discriminate between individuals, the tendency is to select items which measure reasoning, inference, or ability to think. Therefore, the results from these reading tests are artifactually contaminated, usually to an unknown degree, by individual differences in reasoning ability (see Carver, 1973). Lohnes and Gray (1972) argued that in the "...U.S.O.E. Reading Studies measurements were overwhelmingly saturated with general intelligence / p. 466/." If one is studying the relationship between reasoning ability and reading, one should not choose a reading test which had been inadvertently designed to reflect reasoning ability.

Recently, a test has been developed which seems to overcome both of the above problems (Carver, 1977a). The test is called the National Reading Standards (NRS). This test was designed to measure the most difficult material that an individual can accurately read in a reasonable length of time. The passages on the test were selected to reflect a scale of difficulty and the items on the test were objectively chosen using an algorithm to reflect the difficulty of the passages (see Carver 1975). The test reflects reading ability in a manner that is relatively independent of the subjective judgments of the test constructor and not artifactually contaminated by reasoning ability. The NRS provides a scale which measures reading ability in grade level units so that a grade level score of 5 may be interpreted as indicating that the individual can read and understand approximately 50% of the reading material that is written at the fifth grade level of difficulty.

In the development of the NRS, described above, the Raven was also administered to the same students as was the NRS, from Grade 2 to Grade 12. Using these data, the Raven was rescaled to provide an interval scale (see Appendix A). Also, a grade equivalent score was developed so that a Raven raw score can be converted into an expected reading grade level (see Appendix B). The Type I poor readers can be operationally defined as those who have an actual reading ability score, from the NRS, which is approximately equal to their expected reading ability score, from the Raven. Likewise, the Type II poor readers are those who have an actual reading ability score which is lower than their expected reading ability score. The technique used for converting the scores from the Raven into expected reading ability scores is explained in detail in Appendix B.

READING PRACTICE

Practice seems to be one of the most important variables affecting complex cognitive activities. Simon and Chase (1973) studied chess players and they estimated that a master chess player has spent around 10,000 to 50,000 hours staring at chess positions while a Class A players has only spent 1,000 to 5,000 hours doing this. They ask, how does one become a master in the first place, and then they state:

> The answer is <u>practice</u>--thousands of hours of practice....what is needed is to build up in long-term memory a vast repertoire of patterns and associated plausible moves....such a learning process takes time--years--to build up the thousands of familiar chunks needed for master level chess /p. 403/.

This idea about the importance of practice has also been applied to reading with great fervor by Smith (1973). He states:

> All proficient readers have acquired an implicit knowledge of how to read, but this knowledge has been developed through the practice of reading, not through anything that is taught at school ....Learning to read is not a matter of mastering rules. Children learn to read by reading /p. 184/.

Smith elaborates upon this point in another publication (1971):

> What are the circumstances in which the skill of immediate meaning identification is acquired?...I can sum up the answer in a single word, "experience" (or to use a slightly more traditional term, "practice"). Learning to read is akin to any other skill; there are perhaps some specialized exercises that one can undertake to iron out particular difficulties, but there is no substitute for engaging in the activity itself. Reading involves looking for significant differences in the visual configuration to eliminate alternatives, and knowledge can be acquired of what differences are significant only through experience. This knowledge cannot be taught, it has to be acquired; the major contributions that the teacher can make are to provide information, feedback, and encouragement.

The theoretical rationale seems straight forward. If we wish to help poor readers read better, we should get them to practice the act of reading, i.e., we should provide reading training that forces the students to engage in reading.

Practicing reading is seemingly a simple concept but it is not simple to actually manipulate reading practice in an experiment. If an individual is handed a reading passage and asked to read it, the nominal stimulus is straight forward, but the effective stimulus is largely unknown. Anderson (1970) quite simply stated the problem as follows: "One cannot be sure what a student is doing when he is looking at the pages of a textbook /p. 349/." A solution to this problem is now available through the use of programmed prose materials (see Carver, 1975). An example of programmed prose is presented in Figure 1. Notice that an individual must choose which words belong in the sentences of a passage. This task forces an individual to actively engage in reading and it can provide feedback to the experimenter regarding the reader's accuracy. When the programmed prose task is implemented with a computer, it is also possible to give the student immediate feedback; the correct answer to each item can be given immediately following the student's choice. A student can be required to read a passage to a certain criterion of mastery, such as 100% correct. With this technique, it is possible to meaningfully and accurately measure the number of tasks mastered, mentioned earlier in the task-mastery hypothesis. Also, by recording the time required for mastery of each task, the total amount of time spent reading can also be measured. Thus, the practice theory discussed

This is ☐ country / ☐ our Post Office. It is ☐ there / ☐ in our city.
Many people ☐ post / ☐ work here. There is a ☐ Post / ☐ Be Office in
every city ☐ is / ☐ in our country. And Post ☐ Good / ☐ Offices in
every country in ☐ the / ☐ city world.

A Post Office ☐ helper / ☐ in must be honest. He ☐ handles / ☐ must
be a good worker. ☐ Every / ☐ A Post Office helper handles
☐ lots / ☐ and of mail. A Post ☐ animals / ☐ office helper handles lots of
☐ and. / ☐ money.

The Post Office sends ☐ worker / ☐ letters and packages, magazines,
and ☐ helper / ☐ newspapers all over the world. ☐ It / ☐ Helper sends small
animals and ☐ plants, / ☐ all, too. It sends money ☐ for / ☐ office us. It
saves money ☐ sends / ☐ for us. It puts money ☐ to / ☐ and work for us
too.

Fig. 1. An example of a programmed prose passage.
(Taken from Carver, 1975).

above can be implemented using a computer to administer the programmed prose technique.

There is one further matter to be discussed relevant to practice. It would seem that the practice should take place using a level of material difficulty that was near or at the level of reading ability of the individual. Using conventional reading ability tests and conventional reading difficulty formulas, it is not as simple to operationalize this idea as it appears to be. This is because conventional reading ability tests have not been designed to measure ability in terms of the level of material difficulty that an individual can read. The NRS test, described earlier, has a ready-made solution to this problem since material difficulty and individual ability are measured along the same scale. Thus, the difficulty of the material used to train readers, the ability of the readers, and the expected level of reading ability are all measured along a measurement scale that is in fact equivalent instead of being superfically equivalent. This state of affairs makes research much more precise and internally consistent.

EXISTING DATA

Any time a reading test and an intelligence test are given to the same group, it can be predicted that there will be a high correlation between the two. For example, most of the correlations between the STEP Reading Test and the SCAT Test (an intelligence test) are reported to be above .80, according to the manual for the STEP test. It is for this reason that R. L. Thorndike concluded that poor readers are poor thinkers. Yet, there is a lack of good evidence relevant to what happens when poor readers practice reading. There are a number of so called expentancy formulas in reading which use a measure of intelligence to get an expected grade level of reading ability. However, these formulas are usually validated by correlating them with actual grade levels (e.g., see Dore-Boyce, Misner, & McGuire, 1975). What is needed is experimental evidence regarding whether or not those individuals who are actually reading well below their expected level do in fact gain more from high quality instruction than those who are not reading below their expected level.

OVERVIEW OF METHOD

Type I and Type II high school students who read poorly were given reading training. The training consisted of reading practice using a computer terminal to administer the programmed prose technique described earlier. Part of the original research hypothesis was that the Type II readers would become better readers as a result of such practice since it was hypothesized that the primary reason they read poorly was because they had spent little time reading. The other part of the research hypothesis was that the Type II readers would gain more as a result of the training than the Type I because the reason the Type I read poorly was not because of a lack of practice but because of a lower intellectual capacity.

All the students for the research were selected because they read equally poorly. However, one-half of those selected had high scores on the Raven test, mentioned earlier, and the other half had low scores on this test. The Low Raven group therefore represented the Type I poor readers and the High Raven group represented the Type II poor readers.

The training was administered using a PLATO IV computer terminal which provided immediate feedback to the students regarding the correctness of their responses and automatically monitored progress; a student could not progress to a subsequent reading passage until the present one had been mastered to a 100% correct criterion. The primary advantage of the PLATO terminal was that it provided objective and reliable control of the practice. The primary disadvantage was that only one student could be given the reading training at any one point in time. Thus, only a few students could be given the training.

The research was conducted in three studies. Study I involved 6 students for an 8 week summer session, July and August. Study II involved 6 students for a semester, September to December. Study III involved 6 students for another semester, January to April.

STUDY I

INTRODUCTION

Although there was pilot data collected prior to Study I, this study was still in many respects a pilot study itself. It involved the first group of poor reading, high school students who were administered the reading training for an extended period of time.

METHOD

Selection of Subjects. The six subjects involved in Study I were the end result of a lengthy selection procedure.

Two local high schools were asked to cooperate in the project. A request was made to each school to select students in grade 10 with reading levels approximately grade 3 through grade 7. In one school, School A, the reading coordinators selected such tenth grade students from their own classes and they also asked the teachers in the English classes to submit names. This combination of efforts resulted in a list of 64 individuals who were invited to be tested. In the other school, School B, the counseling department used teacher and counselor records and also standardized test scores to select students from grades 9, 10, and 11 who probably were reading around grade levels, 3 to 7. This effort resulted in a list of 100 individuals who were invited to be tested.

In both schools, the Paragraphs Test and the Vocabulary Test from the Literacy Assessment Battery (Sticht & Beck, 1976) were administered to the selected students. The test instructions were tape recorded for standardization and convenience in administration. The entire testing required only about 50 minutes.

In School A, the tests were administered to 48 students in four different groups. In School B, the tests were administered to 56 students in one large group.

There is a reading and auding form of both the Paragraphs Test and the Vocabulary Test so there were a total of four different raw scores. Using tables presented in the test manual, each raw score was converted into a grade equivalent score. From the total of 104 students tested in both schools, the 40 students who had one grade level score less than 6.5 on one of the two reading test were invited to come to the research site for additional testing.

From the 40 invited for additional testing, 24 responded and participated. In two testing sessions, these 24 students were administered a battery of tests--the National Reading Standards, Form 2A, the Raven Progressive Matrices Test, the National Reading Standards, Form 3A, and the Vocabulary section of Survey D, Form 1, Gates-McGinitie Reading Test. They were paid $5.00 for two hours of testing.

The two forms of the National Reading Standards jointly provided a single measure of reading ability in grade level terms. A rescoring system for the Raven Progressive Matrices Test (see Appendix A and Appendix B) provides a measure of intellectual ability in grade level terms. The score on the Gates-McGinitie also provides a measure of reading vocabulary in grade equivalents.

The final 6 students selected were all just finishing their junior year of high school -- four were boys and two were girls. These six students all were reading around the fifth grade level. Their grade equivalents in reading ability were between 4 and 6 as measured by either the National Reading Standards or the Gates-McGinitie. Three of the students had high Raven scores, i.e., grade equivalents of 7-9, and three of the students had low Raven scores, grade equivalents of 2-4. Thus, the Raven grade equivalent scores from 3 of the students indicated that they were reading at a grade level that was approximately commensurate with their grade level of reasoning ability (Low Raven Group), and the Raven scores from the other three students indicated that their grade level of reasoning ability level was higher than their grade level of reading ability (High Raven Group).

Procedures. The students came for training sessions that lasted two hours, five days a week for 8 weeks. The first session started at 8:00 am and the last session ended at 8:00 pm. Each student came every day at the same time period during the day.

The sessions involved interaction with the PLATO computer (Bitzer, Sherwood, & Teczar, 1973) using programs developed especially for this reading training project. There were two programs--the Measuring Reading Efficiency (MRE) program, and the Programmed Prose program (PP). These programs will be described in a later section. The MRE program was administered as pre and post tests for the PP program; this measurement program was not successful and was discontinued after Study I.

Passages. The passages used as reading material for the research were selected from the set of 330, 100-word passages studied by Bormuth (1969). The passages were selected from actual curriculum materials used in schools from grade 1 through college. The readility of these passages in grade equivalents has been measured by the Rauding Scale (Carver, 1975-76). For the present research, ten passages were randomly selected from each set of passages at each of grade levels, 2, 3, 4, 5, 6, 7, 8, and 9 (on the Rauding Scale), giving a total 80 passages.

Measuring Reading Efficiency Program. This experimental testing program was tried out for the first time with these subjects. Each set of 10 passages at each grade level was presented for different periods of time ranging from a very short length of time, 5 seconds, to a relatively long length of time, 2 minutes. After each passage had been presented, there were 10 test items presented one at a time. Five of the items were paraphrases of the information presented in each 1/5 of the passage. When a student was presented one of these items, the student should have depressed a key indicating a "yes" response to the question, Does this sentence say about the same thing as something you read in the passage? The other 5 items were "No" items, i.e., they were not necessarily wrong or untrue items but they did not represent information that was contained in the passage that had just been presented. There was a No item written to correspond to each Yes item, i.e., two items for each part of the passage, Part 1, Part 2, Part 3, Part 4, and Part 5. The 10 items were presented in order from Part 1 to Part 5 and Part 1 to Part 5 again with the Yes or No item for each part randomly determined.

This type of test turned out to be exceedingly difficult for these Ss, even for grade 2 passages. The items were so difficult that the data were not useful enough to present.

Programmed Prose Program. An example of the output from this program is presented in Fig. 2. Every fifth word was an item choice. In this example, the

This is our Post Office. It is in our city. Many people work here There is a Post Office in every city in our country. And Post Offices in every country in the world.

A Post Office helper must be honest. He must be a good worker. A Post Office helper handles lots of mail. A Post Office helper handles lots of money.

The Post Office sends letters and packages, magazines, and newspapers all over the world. It [of*sends] small animals and plants, [handles*too]. It saves money for [us*plants]. It puts money to [work*money] for us, too.

Fig. 2. An example of the programmed prose task as implemented by the PLATO IV computer terminal.

student has already responded to the first 16 items; the remaining four items are represented by the four boxes enclosing the four pairs of words. The student's task was to choose either the right or left word as best fitting into the sentence by pushing the appropriate key on the right or left side of the keyboard in front of them. The alternative wrong words were choosen by an algorithm, described in detail elsewhere (Carver, 1975). When the student chose the correct word for an item, the incorrect word and the box simply disappeared. When the student chose the incorrect words, the word INCORRECT flashed briefly immediately below the passage; the incorrect word and the enclosing box disappeared as soon as INCORRECT was through flashing. After a passage had been completed, a feedback page appeared with the following information given: number right, total number of items, time taken in min., and Rate of Good Reading (RGR) score. The RGR score is a type of reading efficiency score which is computed from a formula which combines the accuracy, rate, and grade level of difficulty of the passage. The higher the accuracy, the higher the RGR score. The faster the rate, i.e., the lower the amount of time taken to complete the passage, the higher the RGR score. The more difficult the passage, the higher the RGR score. The formula itself was a modification of the reading efficiency formula presented in the manual for the National Reading Standards. That formula adjusted for the difficulty of the material by use of average word length. The present formula uses the equations given by Carver (1976; 1977a) for predicting average word length from grade level of difficulty. The end result is the following equation:

$$RGR = \frac{1.468\ (2.4R-1.2T)(.0787G+3.861)}{t} + 136.7 \qquad (1)$$

where RGR is the Rate of Good Reading,
R is the number of right or correct items,
T is the total number of items for the passage (20 items),
G is the grade level of the passage as rated by the Rauding Scale,
and t is the time in minutes spent working on the items.

Notice that the lowest possible RGR score, at an infinitely large amount of time or an infinitely low rate, is 136.7 words per minute. Since the grade level of difficulty of the passages, G, varied only from 2 to 9, the average word length term in Equation 1 varied only from 4.0184 to 4.5693 character spaces per word (see Carver,1976). Thus, the effect of the grade level difficulty upon RGR varied only 13.7% from the grade 2 to grade 9 training materials.

The students were instructed that the first time they saw a passage they should treat it as a test: "You should try to get the items right; you should try to work as fast as you can; you should try to get your RGR score as high as possible." After the first trial on the passage, they were then instructed to take their time and try to get all of the items correct. They then kept working on a passage until they got all the items correct. Each new trial on a passage presented the same passage but new items. There were five different possible sets of correct items, i.e., the first could start with the first word, second word, third word, fourth word, or fifth word. Even if two trials on a passage involved the same set of correct answers, the wrong answers would be different since for each new trial the incorrect alternatives were randomly selected, with certain restrictions (see Carver, 1975), from the surrounding text. After the student got all the items correct (there were 20 items), then the student took two tests on the passage prior to going onto the next passage. The instructions for the tests were exactly the same as on the first trial and the items for the tests were selected according to the same algorithms used on all the other trials. These two tests also gave the student practice in going fast immediately prior to taking the test upon the first presentation of a new passage.

The students proceeded through the 80 passages in order of difficulty from grade 2 through grade 9 with the order of the 10 passages within each grade level randomly determined, initially, but constant and the same for all students.

After each trial, the student recorded the data from the feedback page onto a data sheet. These data sheets provided the data that were subsequently analyzed. These data were also recorded by the computer and a hard copy printout allowed the accuracy of the student to be verified. No discrepancies were noted. When the data had been recorded, the student pressed the space bar on the keyboard and the computer started preparing the next passage to be presented. This required about 20 to 30 sec. depending upon the load upon the computer at that particular time. The time required to complete a passage ranged from about 1.5 min. to 3.0 min. depending upon the student's rate of working. The intertrial interval including data recording time and processing time for the computer was around .5 to 1.0 min.

Order of Presentation of MRE and PP. The Grade 2 and Grade 3 sets of 10 passages were considered as practice by the experimenter although the students were not told this. The MRE testing on Grade 2 and Grade 3 passages preceeded the PP activity on the Grade 2 and Grade 3 passages. The MRE testing on subsequent passages took up most of the initial sessions but PP on Grade 2 and 3 was also initiated during the initial sessions. The grade 4-6 passages were treated as one block. All of the MRE testing on grade 4-6 passages had been completed prior to the PP activity on the grade 4-6 passages. When the PP activity on the grade 4-6 passages was completed, then the MRE post testing on the grade 4-6 passages and the MRE pretesting on the grade 7-9 passages was completed prior to resumming the PP activity. When the PP activity was completed on the grade 7-9 passages the MRE testing on that block of passages was initiated.

The very first passage presented to the students, i.e., the first Grade 2 passage, was designated as a practice passage. This practice passage was presented prior to beginning the PP each session. This procedure allowed a warmup to a constant criterion of performance prior to the continuation of the PP activity.

Data Analysis. The primary variable was the RGR score on the first trial which was a test trial where the student had been instructed to try to get the RGR score as high as possible by going fast without making very many mistakes. The RGR score was probably the best index of whether the students are improving their general reading ability. This measure combines rate and accuracy so any improvement in either of these two compoments of reading efficiency should be reflected in this score. This first trial RGR will be called the PRETEST RGR.

The other two variables were the number of PASSAGES MASTERED and the amount of time spent mastering passages, i.e., TOTAL TIME. The PASSAGES MASTERED variable was scaled, for convenience, as 2, 3, 4, 5, 6, 7, 8, and 9 with these 8 numbers from 2 to 9 representing the 10 passages at that particular grade level of difficulty, e.g., "2" designates the 10 passages at grade 2 difficulty.

For data analysis purposes, the 10 passages at each grade level were broken into two subsets of 5. The median RGR score for each consecutive subset of 5 passages was the primary unit of analysis. Thus, there were two blocks of five passages at each grade level since there were 10 passages at each grade level altogether. The mean of these two medians for each level of passage difficulty, e.g. 2, provided an indicator of the typical performance of the individual for that set of 10 passages; the indicator of typical performance for each of the two groups, i.e., Low Raven and High Raven, was obtained by calculating the mean of these three means for the three individuals in each group.

The TOTAL TIME variable included the first trial test, all trials to mastery, and the two post tests. The total time for the first block of 10 passages, for example, was simply the sum of the total times for each of the first 10 passages. The time for the second block was the sum of the times for the first block and the second block, i.e., the total time spent up through the second block. Thus, the TOTAL TIME variable was the total time spent on the reading tasks from the beginning up through each successive block of 10 passages.

RESULTS AND DISCUSSION

Fig. 3 contains the PRETEST RGR scores as a function of PASSAGES MASTERED for the Low Raven and High Raven groups.* For example, the RGR value for the High Raven group on the grade 2 passages (RGR=197) was the mean of the two median values representing the first and second block of 5 passages for all three students in the High Raven group. Notice that the three students with High Raven scores are not distinguishable as a group from the students with Low Raven scores. The Low Raven group did start out somewhat lower but as more passages had been mastered, the Low Raven group improved to a point where their RGR scores were slightly higher than the High Raven group. This result is in accordance with the research hypothesis which held that there would be no differences between the Low Raven group and the High Raven group when the number of passages mastered was held constant.

---

*Note that the scale in Fig. 3 seems disproportionate. This scale was selected because it is consistent with the remainder of the data to be presented in Study II and Study III.

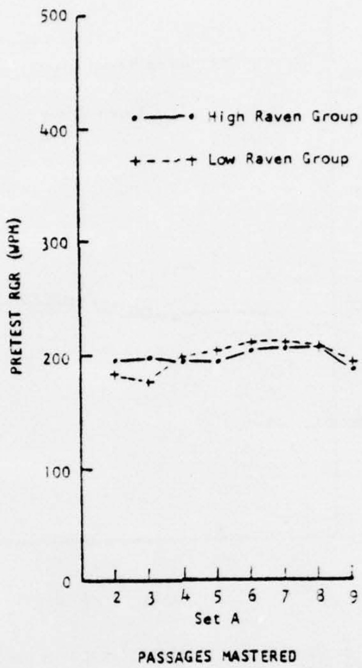


Fig. 3. Study 1. PRETEST RGR, in Wpm, as a function of passages mastered for the Low and High Raven groups.

In Fig. 4, the PRETEST RGR scores are presented as a function the amount of TOTAL TIME. These data do not present results that are significantly different from those in Fig. 3. The High Raven group is not distinguishable from the Low Raven group. This result is not in accordance with the research hypothesis. When time was held constant, it was hypothesized that the High Raven group would gain more as compared to the Low Raven group. There is no evidence to support this hypothesis in Fig. 4. The individuals with supposedly lower reasoning ability seemed to gain just as much in a fixed amount of time as the group of individuals with supposedly higher reasoning ability. There was no evidence in Study I to support the research hypothesis regarding Type I and Type II poor readers.

It is possible that the Raven scores for this group were not reliable. It could be that some of the students in the Low Raven group simply did not try their best and were therefore categorized erroneously. Eysenck (1966) states that the two main personality factors which influence performance on intelligece tests "...may be called by their popular names, carelessness and lack of persistence /p. 16/." Thus, it could be that some of the students in the Low Raven group were impulsive, i.e., refused to persist and carelessly answered with the first answer that came to mind. Before rejecting the research hypothesis outlined at the outset, it seemed prudent to focus on the reliability and validity of the

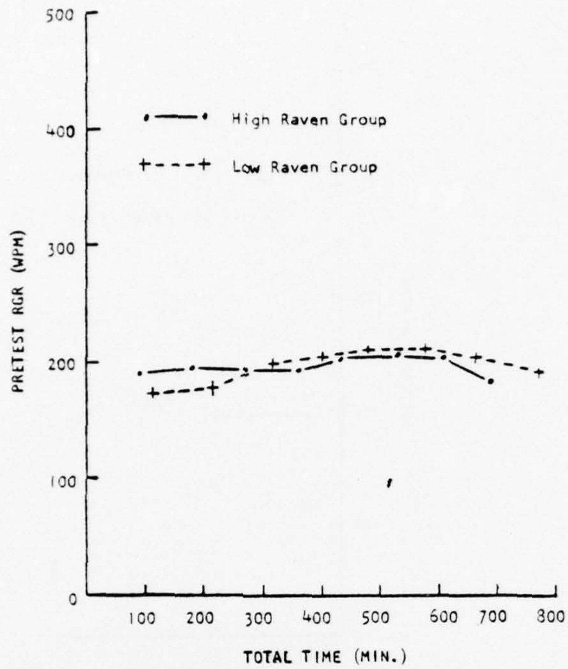


Fig. 4. Study I. PRETEST RGR, in Wpm, as a function of TOTAL TIME, in min., for the Low and High Raven groups.

Raven test, and to modify the experimental procedures in other ways to see if these results were replicable.

STUDY II

INTRODUCTION

This study was simlar in most ways to Study I; however, there were several important differences.

An attempt was made to more reliably measure reasoning ability with the Raven. The method of administering this test in the selection process was changed in a manner that was designed to elicit greater attention from the students. Also, the Wechsler Intelligence Scale for Children was administered to the students who were finally selected to provide another measure of intellectual ability.

The MRE testing was eliminated and the number of passages that were to be mastered was increased from 80 to 160, to provide more reading training or practice. By increasing the amount of training time, any differences that in fact exist should become more pronounced.

There were other differences between Study I and Study II which will be explained in the method section.

METHOD

Selection of Subjects. Again, the six subjects involved in Study II were the end result of an elaborate selection procedure. The procedure was similar to what was used in Study I but deviated in certain ways as will be described below.

In School A, the reading coordinator positions had been eliminated over the summer and the two reading coordinators had been assigned to teach English classes. The initial testing involved the 99 ninth and tenth grade students who attended the first four English classes taught by one of the former reading coordinators in School A. Form 3A of the National Reading Standards was administered to these classes. Those students who scored at grade level 7 or higher on this test were not invited for further testing. After this group had been eliminated, the students with the remaining highest 50 scores were invited for an additional testing session of 2 hours duration at the research site. This testing was conducted after school hours in three, 2-hour sessions, and the students were paid $5.00 for their particpation. A total of 28 students participated in this phase of the testing. Each student was administered the National Reading Standards, Form 2A, the Raven Progressive Matrices Test, and the Vocabulary section of Survey D, Form 1, of the Gates-McGinitie Reading Test. The preceeding tests were also administered in Study I. However, in Study II, the Raven test was administered in a manner that deviated from the standarized procedures. After the test had been administered once in the standard manner, it was scored and the students were asked to take it again toward the end of the same 2-hour testing session. The second time, they were given an answer sheet which nad the ones they had missed the first time marked with an X. Thus, they could see how many they got right and which items they got wrong but they had no clue as to which specific alternatives were right or wrong. This information by itself should not have much of an effect upon the scores because the Raven has 6 alternative answers for the first 24 items and 8 alternative answers for the remaining 36 items.

The six subjects finally selected were all reading at the fourth, fifth or sixth grade levels according to both the NRS test and the Gates-McGinitie

Vocabulary test. Three of the subjects had high Raven scores, 8-9 (High Raven group), and three had Low Raven scores, 3-5 (Low Raven group), as measured by the second administration of the Raven. The six subjects were all starting either their ninth or tenth grade in school--three were boys and three were girls.

Procedures. Each student came for a two-hour session, three days a week. The sessions were scheduled from 3:30 to 9:30 pm, daily and from 12:00 to 6:00 pm. on Saturday. The subjects came at the same time periods each week--three came Monday, Wednesday, and Friday, and three came Tuesday, Thursday, and Saturday.

As mentioned earlier, there were no MRE activities in Study II. The programmed prose activity for each passage was exactly the same as was outlined in Study I but there were changes in the overall PP treatment. In Study I, the experimenter considered the Grade 2 and Grade 3 sets of 10 passages as practice. In Study II, there was no practice except for the first passage--the practice passage; there was no practice from either the subject's standpoint or the experimenter's standpoint. The first 80 passages, passage Set A, were exactly the same as in Study I. The next 80 passages, Passage Set B, paralleled the first set and were selected the same way as the first set; no overlap between sets was allowed. When a subject finished the last passage in Set A, i.e., a passage at the ninth grade level, then the subject was started immediately on the first passage on Set B, i.e., a passage at the second grade level. Again, the student always started out each session with the practice passage.

The fastest student in the group completed the 80 passages in 17 sessions and the slowest student took 34 sessions. When the faster students finished they were asked to start all over again at the beginning to keep them participating in the project. All students completed at least 28 sessions.

Wechsler Testing. The Wechsler Intelligence Scale for Children (WISC) was administered to each of the 6 students toward the end of their first month of participation. It was administered by a research assistant on the project.

Data Analyses. The data analysis procedures were similar to those in Study I but were much more intensive. Although the posttesting on each passage acted primarily as a warmup for the pretesting on the subsequent passage, the data from the posttesting also reflects gain in efficiency of performance as a result of practice. Similarly, the practice passage that was administered at the beginning of each 2 hours session was presented primarily as a warmup but the RGR data from this passage also reflects performance gain. Therefore, the higher of the two posttest RGR values for both the regular passages and the practice passage were analyzed in Study II; this variable was called the POSTTEST HIGHER RGR.

Also in this study, the components of certain variables were anlayzed. The time and accuracy components of the first trial RGR was analyzed, i.e., the PRETEST TIME variable and the PERCENT ACCURACY variable. The PERCENT ACCURACY variable was actually a corrected value, i.e., using the correction for guessing formula--rights minus wrongs. In addition to analyzing the total time taken on each passage, the number of trials required to reach mastery was also analyzed. This latter variable was called TRIALS TO MASTERY.

GROUP RESULTS AND DISCUSSION

In Fig. 5, the POSTTEST HIGHER RGR values on the initial practice passage

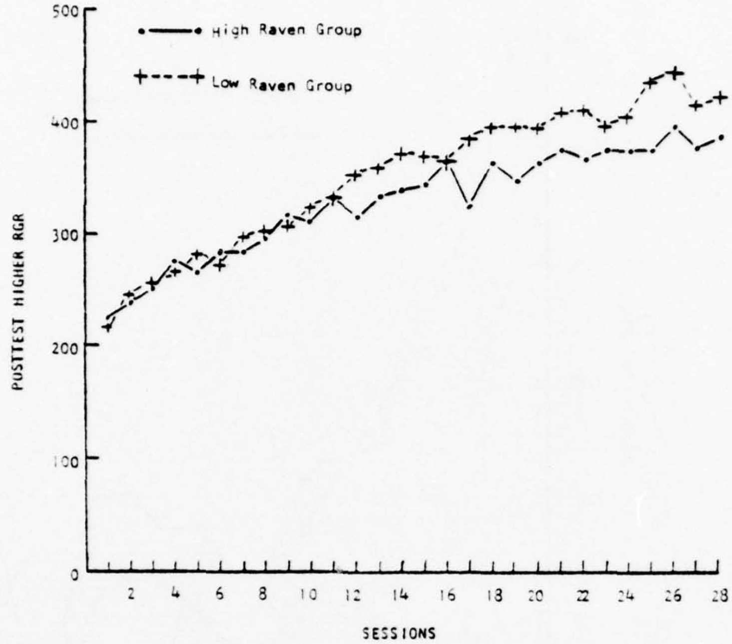


Fig. 5. Study II. POSTTEST HIGHER RGR on the initial practice passage as a function of sessions completed for the Low and High Raven groups.

for each session have been presented for the Low Raven and High Raven groups. On the average, these students started at around 220 words per minute after mastery of the practice passage and steadily improved to about 400 words per min. Notice that both groups start out almost exactly equal (220 and 221) but the Low Raven group consistently does better than the High Raven group from about session 12 on. These data indicate that both groups consistently improved the speed at which they could complete this practice passage. These data also indicate that these groups have an ultimate potential of around 400 words per minute for RGR performance. This means that any RGR scores lower than this reflects the amount of thinking or reasoning time required to process a less familar passage as opposed to a ceiling imposed by the equipment or the simple reaction time of the student.

Fig. 6 contains PRETEST TIME as a function of PASSAGES MASTERED. The values plotted for each grade level of difficulty were the end result of the same type of averaging procedure as was explained in Study I. In Study II there were two sets of passages at each grade level and the second set, Set B, followed the first set, Set A. Notice that the time required to complete the pretest was over 2.00 minutes at the beginning and decreased to about 1.40 min. at the end of the 160 passages that were mastered. The Low Raven group started out considerably slower than the High Raven group, but took less time in most cases on the Set B passage.

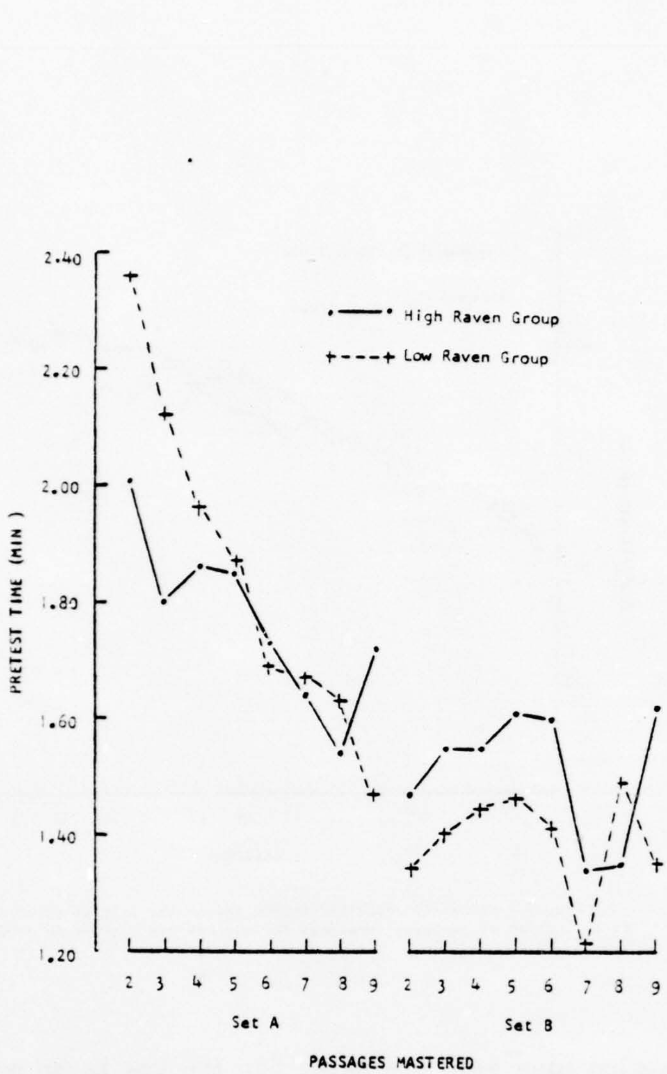


Fig. 6. Study II. PRETEST TIME, min., as a function of PASSAGES MASTERED for the Low and High Raven Groups.

Fig. 7 contains PERCENT ACCURACY as a function of PASSAGES MASTERED. As mentioned earlier, the percent accuracy score is a score that has been corrected for guessing; the number of items answered wrong was subtracted from the number of items answered right and the result was divided by the total number of items and multipled by 100. For both the High Raven group and the Low Raven group, the scores tend to decrease as the level of difficulty of the passages increases from grade 2 to grade 9. In general, there was little difference between the Low Raven group and the High Raven group. Toward the end of the Set B passages the Low Raven group seemed to score about 5 to 10 percent higher than the High Raven group.

Fig. 8 contains the average number of trials to reach mastery, TRIALS TO MASTERY, as a function of PASSAGES MASTERED. The pretest was counted as the first trial so if the students had all reached mastery on the pretest, the average number of trials to mastery would have been its lowest possible value, 1.0. The number of trials to mastery tended to increase as the difficulty of the passages increased from grade 2 to grade 9. There appeared to be no consistent difference between the High Raven group and the Low Raven group. The Set B passages tended to be mastered faster than the Set A passages, thus re-

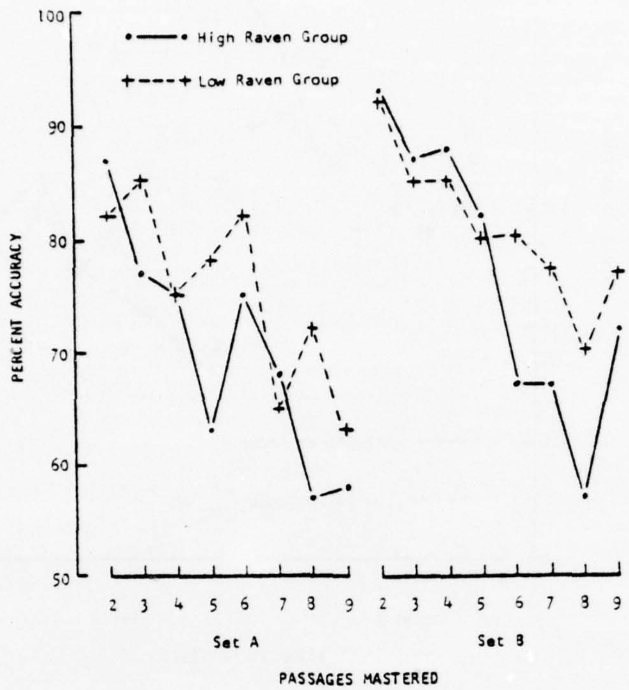


Fig. 7. Study II. PERCENT ACCURACY on the pretest as a function of PASSAGES MASTERED for the Low and High Raven groups.

flecting a learning effect. On the Set B Passages, about $1\frac{1}{2}$ trials were required on the grade 2 and 3 passages and about 3 trials were required on the grade 7-9 passages. These data, taken with the percent accuracy data presented in Fig. 7, suggest that the difficulty of the training material was in fact being manipulated by these passages which were originally designated as grade 2 to grade 9 by the Rauding Scale.

Fig. 9 contains TOTAL TIME as a function of PASSAGES MASTERED. Notice that the Low Raven group tended to take more time than the High Raven group. However, the difference between the two groups at the end of the Set A passages did not change all through the Set B passages. This means that the rate at which both groups worked on the Set B passages was exactly the same. Thus, the rate of learning difference in favor of the High Raven group evaporated after the Set A passages had been mastered.

Fig. 10 contains the POSTTEST HIGHER RGR as a function of PASSAGES MASTERED. Both groups have higher scores as the number of passages mastered increases, starting with an RGR around 225 for the initial grade 2 passages and increasing to around 325 for the final grade 9 passages. There was no consistent difference between the two groups until the Set B passages was reached where the Low Raven group tended to score higher than the High Raven group. Since these test data reflect performance after 100% mastery has been achieved, increases in RGR is largely a result of decreases in the time required to complete the passage. Thus, these data reflect approximately the same trend as the data from the practice

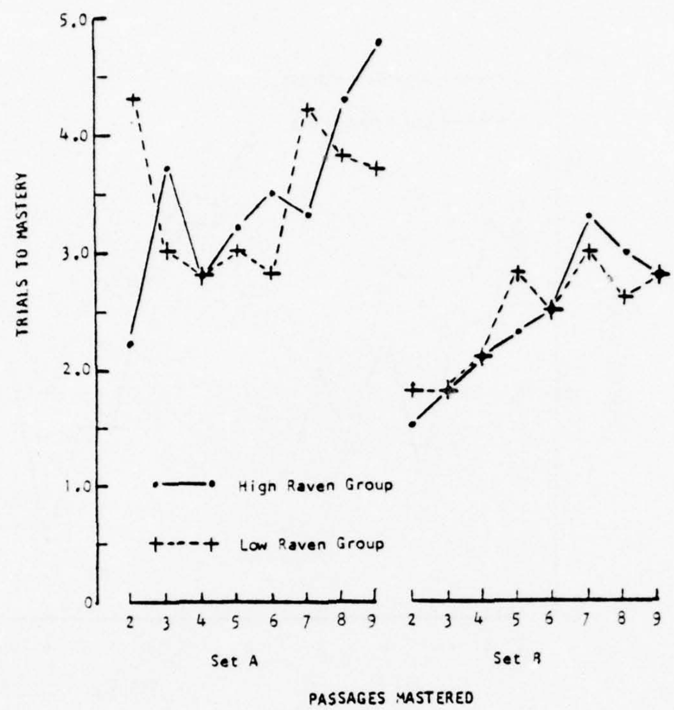


Fig. 8. Study II. TRIALS TO MASTERY as a function of Passages MASTERED for the Low and High Raven groups.

passage presented in Fig. 5. The major difference is that those data do not reach the high values that approach 400, partly because each passage is only mastered once in Fig. 10 whereas the same passage was mastered 28 times in Fig. 5.

Fig. 11 and Fig. 12 contain the primary data from Study II. These data parallel those data presented in Fig. 3 and Fig. 4 for Study I, except the students in Study I did not complete the Set B passages. The research hypothesis was that there would be little or no difference between the Low Ravens and the High Ravens when gain in proficiency was measured with the number of passages mastered held constant, as in Fig. 11. But, there would be a difference in favor of the High Raven group when gain in proficiency was measured with the total amount of time spent learning held constant, as in Fig. 12. This hypothesis received no support from these data since the Low Raven group scored higher in both Fig. 11 and Fig. 12. In spite of the fact that the Low Ravens and the High Ravens started out approximately equal, RGR equal to about 195, the Low Ravens gained to an RGR around 230 after around 900-1100 minutes of effort while the High Ravens had only gained to around 205 after the same amount of time.

In order to determine if this lack of support for the research hypothesis held for another indicant of proficiency, the POSTTEST HIGHER RGR values were also plotted as a function of the total amount of time spent working on the passages. Fig. 13 contains these data. Thus, the data in Fig. 10 and Fig. 13 contain data which are parallel to the data in Fig. 11 and Fig. 12 except the indicant of proficiency is POSTTEST HIGHER RGR instead of PRETEST RGR. THE POSTTEST HIGHER RGR data in Fig. 10 and Fig. 13 also do not support

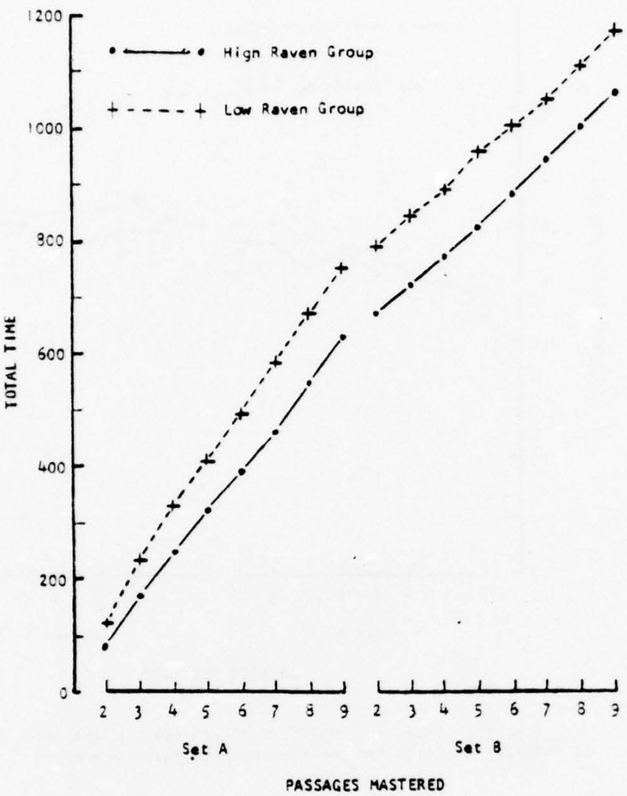


Fig. 9. Study II. TOTAL TIME spent working on the passages as a function of PASSAGES MASTERED for the Low and High Raven groups.

the research hypothesis. These data indicate little or no difference between the RGR scores of the Low and High Raven groups for either the passages mastered or total time variables.

Even though there were no consistent differences between the Low and High Raven groups, there were large differences within the Low Raven group. The fastest and slowest learners of all 6 students were both in this Low Raven group. This was true for PRETEST RGR, POSTTEST HIGHER RGR, and also for the POSTTEST HIGHER RGR on the practice passage. Fig. 14 contains the POSTTEST HIGHER RGR data for the practice passage presented for each session. Although S4 (student 4) started about 55 words per minute higher than S6, S4 was over 195 words per min. higher than S6 by Session 28. Also included in Fig. 14 are the data from 2 research assistants, graduate students, who worked on the project. These two graduate students followed exactly the same procedures as the other 6 students except they completed the Set B passages before the Set A passages and there was an interuption of over one month between the completion of the Set B passages and the starting of the Set A passages.

Notice that the two graduate students, S7 and S8, completed all 160 passages in only 4-5 sessions whereas the other 6 students required around

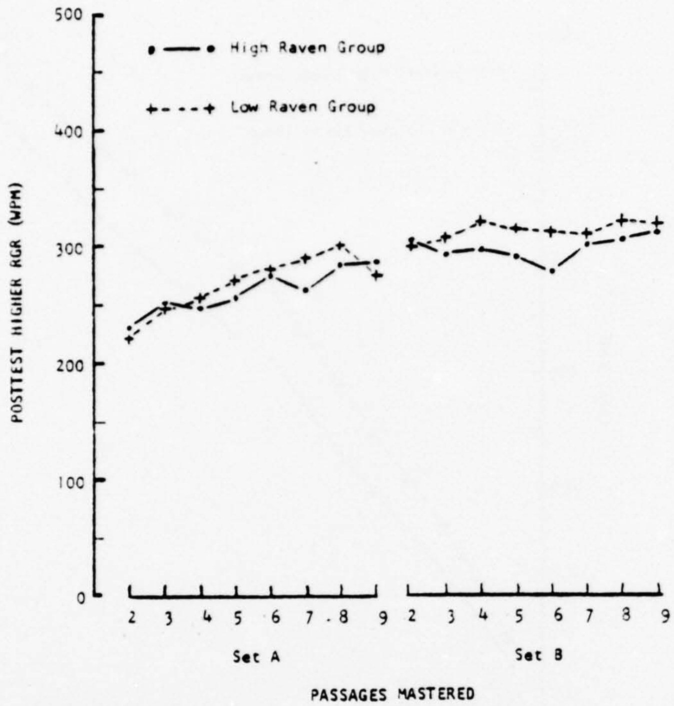


Fig. 10. Study II. POSTTEST HIGHER RGR, in Wpm, as a function of PASSAGES MASTERED for the Low and High Raven groups.

25-30 sessions. These two students had achieved a performance level by the fourth session that was higher than 5 of the other six students had reached in 25 sessions.

The rate of gain per session for the two graduate students was much higher than the other six students even including S1, the fastest of the other six. Although the rate of gain per session varied considerably among the 8 students, the rate of gain per passage mastered was relatively constant across the 8 students; these data are presented in Fig. 15. (Remember that the two graduate students actually worked on the Set B passages prior to the Set A passages.) It can be seen that all 8 curves are approximately parallel indicating that the rate of gain per passage mastered is approximately constant accross all eight students. This type of result was what had been hypothesized for the PRETEST RGR data from the Low and High Raven groups. These data in Fig. 15 indicate that no matter what the level of ability, as indicated by the Raven, everyone tends to gain the same amount when the amount of quality practice is held constant. However, the data in the earlier figure, Fig. 14, indicates that the Raven was not a good predictor of the differences in gain per equal amounts of time spent learning. This is because time is constant in Fig. 14 and one of the students in the Low Raven group turned out to be by far the fastest learner of the six high school students.

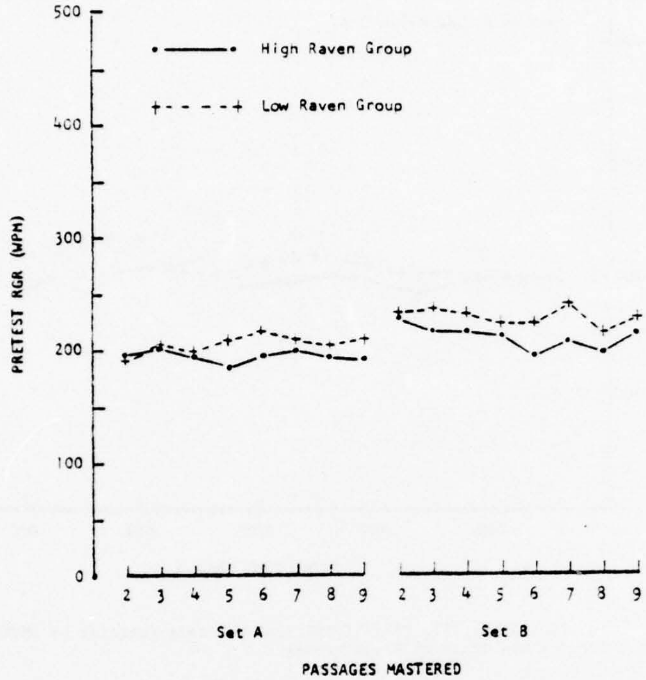


Fig. 11. Study II. PRETEST RGR, in Wpm, as a function of PASSAGES MASTERED for the Low and High Raven groups.

RAVEN RESULTS AND DISCUSSION

Introduction. The results from the three students in the High Raven group tended to be approximately the same in Fig. 14, but the data from three students in the Low Raven group was very disparate. One of these three Low Raven students was a very fast learner, S4, while another one was approximately equal to the High Ravens, S5, and one was very slow, S6. It seemed possible that the theory might be valid, in spite of the data presented, if the Raven data was somehow inaccurate or unreliable. Suppose S4 and S5 happened to belong in the High Raven group but tested out to be in the Low Raven group simply because they failed to do their best on the test. This section will contain evidence relevant to that possibility. Since there was such a lack of homogeneity in the Low Raven group, the results from each student will be presented.

Wechsler Data. Table 1 contains the data from the WISC. The median Verbal IQ of the High Raven group is 79 and the median for the Low Raven group is 76. The lowest Verbal IQ was for S6, the slowest learner as discussed earlier. For the Performance IQ's, the median of the High Raven group was 104 and the median of the Low Raven group was 79. These data seem to replicate the reading test data and the Raven data used to select these students. The reading test data is in parallel with the Verbal IQ data and the Raven test data is in parallel with the Performance IQ data. The Raven grade levels were used to select students who seemed to have high potential for learning to read better. The three High Raven students had a median Performance IQ, 104, which is considerably higher

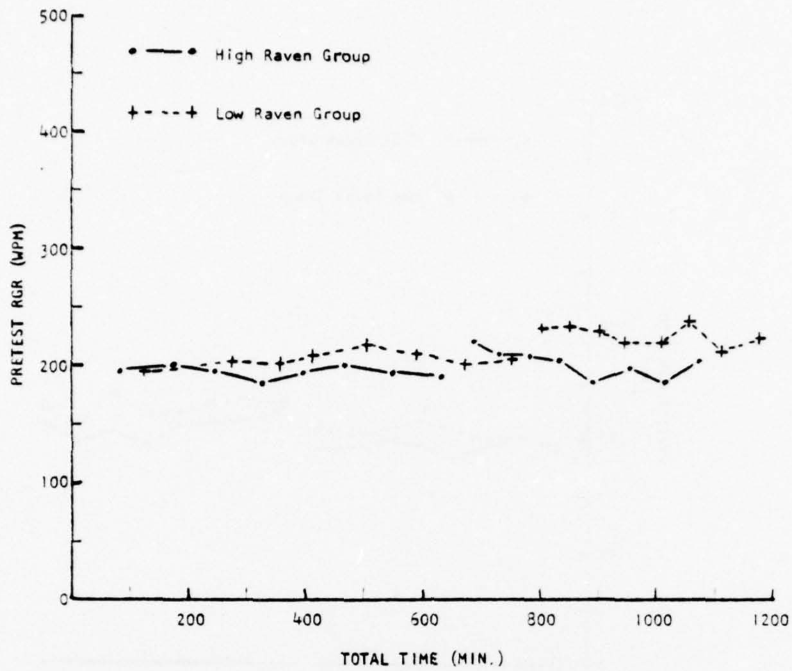


Fig. 12. Study II. PRETEST RGR, in Wpm, as a function of TOTAL TIME, in min., for the Low and High Raven groups.

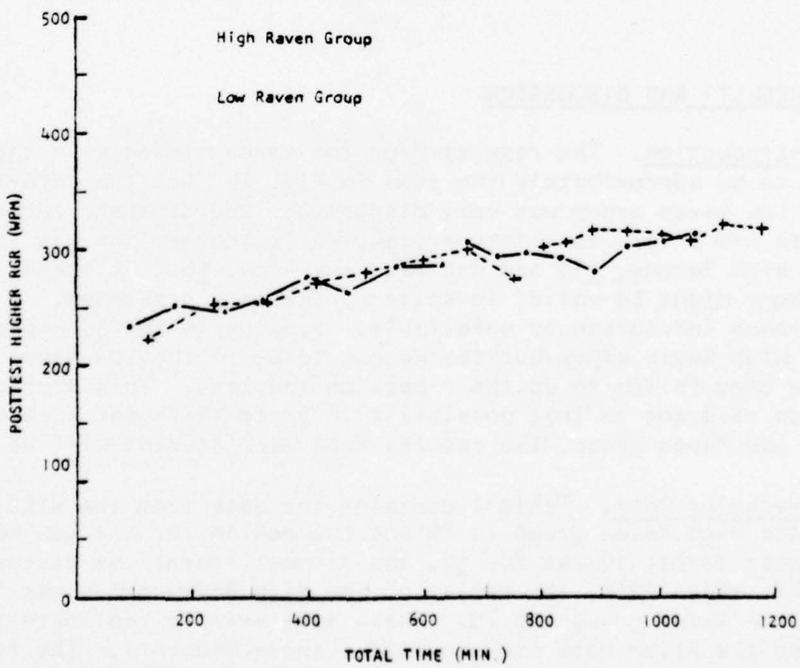


Fig. 13. Study II. POSTTEST HIGHER RGR, in Wpm, as a function of TOTAL TIME, in min., for the Low and High Raven groups.

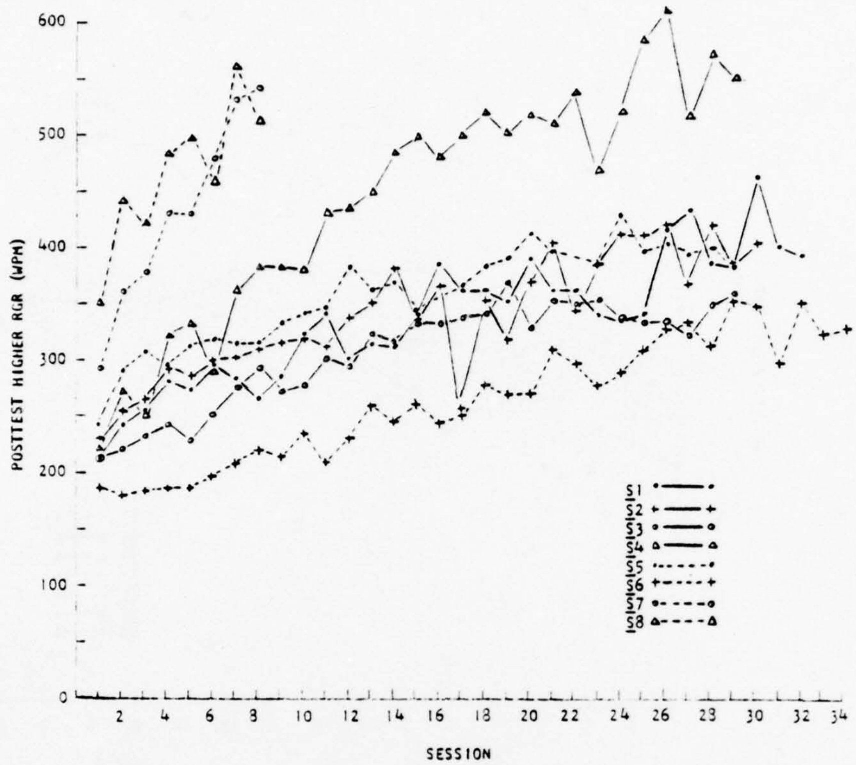


Fig. 14. Study II. POSTTEST HIGHER RGR, in Wpm, for the practice passage as a function of the training session attended for each subject.

than their median Verbal IQ, 79, which also would suggest high potential for this group. The Low Raven group was selected because their Raven grade levels suggested low potential with respect to learning to read better. These three students have a median Performance IQ, 79, which is connensurate with their median Verbal IQ, 76, indicating little potential from a verbal standpoint. Thus, the WISC tends to reinforce the reliability and validity of the Raven rather than suggesting that it was inaccurate.

Raven Data. The hypothesis about the inaccuracy of the Raven was investigated further in spite of the WISC data which supported its reliability and validity. The reason the fast learners in the Low Raven group scored low on the Raven may have been because they did not try their best or because they were impulsive. A small subsiderary data collection effort was initiated toward the end of Study II to investigate these possibilities. The Raven was administered again, individually, under quite different conditions. The student orally stated the answer to each item and was immediately given a dime, 10¢, if the answer was correct. If the student was wrong, the student was asked to try again. To reduce impulsivity, the student was asked to wait one min., as indicated by a stop watch that the student could see, before giving the second answer. The two scores which resulted as well as the scores from the two earlier testings, are presented in Table 2. For ease of interpretation, the grade equivalent scores for each individual have been presented and they have been rounded off to the nearest grade level.

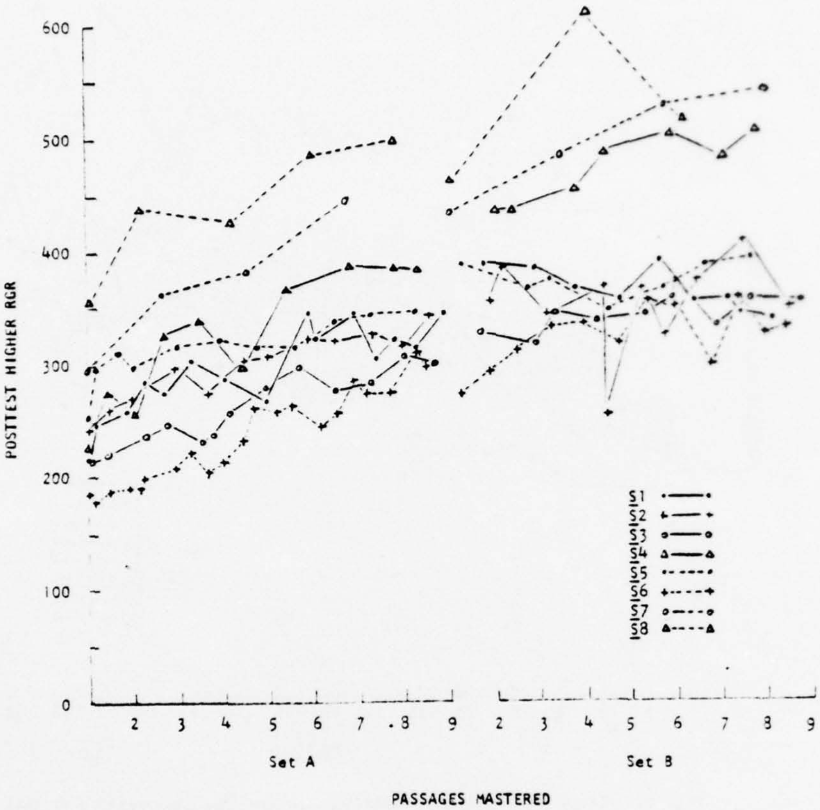


Fig. 15. Study II. POSTTEST HIGHER RGR, in Wpm, for the practice passage as a function of PASSAGES MASTERED for each subject.

Notice in Table 2 that the fastest learner, S4, still made a low Raven score, 5, even when paid 10¢ for each correct answer on the third administration of the test. Thus, lack of motivation did not seem to be a good hypothesis for explaining the low Raven score. However, when the second try results, i.e., the fourth administration of the test, were scored as though they were simply another Raven Test, it can be seen that S4 greatly improved his score (to 9) almost to a point where it equalled the scores in the High Raven group (10,10, & 11). S6 also considerably improved his score as a result of the fourth testing session. S5, however, scored closer to S6 on the Raven while being closer to S4 and the High Raven group in terms of rate of learning.

These data are mixed with respect to explaining the lack of support for the research hypothesis. It seems reasonable that the potential of S4, the fastest learner, was not accurately measured by the Raven because of his impulsivity. It could easily be that this student has an extremely poor strategy for tackling these problems. It seems reasonable that this student would do much better on the Raven than the other Ss if he could be trained to take his time, and systemtically eliminate all wrong alternatives before choosing his answer. Similarly, it would seem that this student would have done much better on the WISC performance scale if he was trained to take his time and

TABLE 1

Verbal and Performance IQ Scores from the WISC for the 6 Students

Group	Student	Verbal	Performance
	1	74	118
High Raven	2	79	104
	3	85	96
	Median	79	104
	4	76	85
Low Raven	5	81	79
	6	56	76
	Median	76	79

TABLE 2

Raven Scores in Grade Equivalents

for the 6 Students in Study II

TEST ADMINISTRATION

S	First	Second	Third	Fourth
1	5	8	9	11
2	6	8	8	10
3	9	9	9	10
4	5	3	5	9
5	4	5	6	7
6	2	3	3	6

do his best. However, as satisfying as this explanation for S4 may be, the data of S5 still was inconsistent. S5 did not make much progress even with the repeated testing. S5 was as fast a learner as the Ss in the High Raven group yet had a Raven score that was closer to S6, the slowest learner.

There is one additional piece of confounding information that needs to be reported. S4 and S5 turned out to have been school friends prior to the experiment. They participated on different days but kept track of each other's progress in order to compete with each other. For example, S4 sometimes did not take the 10 min. break that was allowed at the end of the first hour so that he could work to get ahead of S5. Thus, variation in the motivation variable during training may have confounded the results.

STUDY III

INTRODUCTION

The data and results from Study II had implications for the design and data collection in Study III.

It was decided that the focus in Study III should remain upon comparing High Ravens with Low Ravens to see if the lack of support for the original research hypothesis could be replicated a third time. It was noted in an analysis of the individual data in Study II (not presented earlier) that the student who gained as much or more than anyone on PRETEST RGR was the student with the lowest WISC Verbal and Performance scores. Therefore, it seemed reasonable to attempt to select the next set of 6 students with slightly lower reading ability at the outset. The training procedure may result in gains that are task specific for the betters readers but are more likely to be generalizable to normal reading for the readers at a slightly lower level of reading ability.

METHOD

Selection of Subjects. Of the 28 students in Study II who came for the 2 hr. testing session that was part of the selection procedures, 9 of that group was invited to come again for another 2-hr. testing session. These nine students had reading test scores which were just lower than the 6 students who were selected for Study II. The lowest reading ability score of the 6 Study II subjects, initially, was grade 4, and the reading ability scores of the individuals in this Study III group were grade 4 or grade 3.

From this group of 9 which was invited to particpate and told they would be paid $5.00 for their participation, 7 actually came for the testing.

At this second pretesting session, Form B of the NRS was administered, both Level 1 and Level 2. Then the Raven was administered again. The Ss were reminded that they had taken the Raven twice before, back in September. They were told that the items on answer sheets that they missed the second time they took the test had been marked with an X. They were asked to take the test again, trying to make sure they got the same items right again and trying harder on the ones they missed before.

Six of the seven were selected to participate in Study III--five girls and one boy. Using the student's highest score on any of the forms of the NRS that were administered, these six Ss this time all scored between grade 4 and grade 6. The three students choosen for the High Raven group scored at grade 9 and grade 10 on this third administration of the Raven. The three students choosen for the Low Raven group scored at grade 6 or grade 7.

Procedures. The training procedures were exactly the same as reported in Study II.

Data Analyses. The data analysis procedures were also exactly the same as reported in Study II.

GROUP RESULTS AND DISCUSSION

In Fig. 16, the POSTTEST HIGHER RGR values on the initial practice passage for each session have been presented for the Low and High Raven groups. All students in each group finished at least 25 sessions. These data replicate the results of Study II (presented in Fig. 5) in that there was little difference

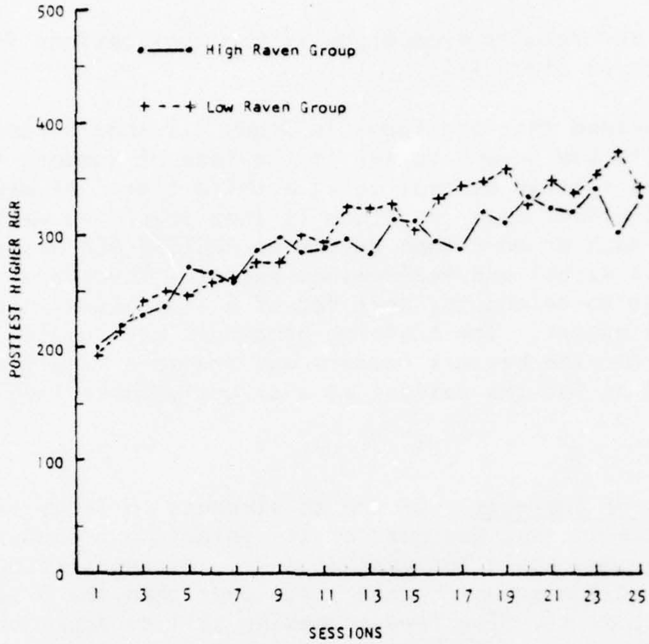


Fig. 16. Study III. POSTTEST HIGHER RGR on the initial practice passage as a function of sessions completed for the Low and High Raven groups.

between the two groups and what difference there was tended to favor the Low Raven group. These students in Study III started slightly lower in session 1 than the Study II students and were also slightly lower at Session 25.

Fig. 17 contains PRETEST TIME as a function of PASSAGES MASTERED. These data tend to parallel those in Study II (see Fig. 6) except the Low Raven group performed consistently faster than the High Raven group from the first to the last of the 160 passages.

Fig. 18 contains PERCENT ACCURACY as a function of PASSAGES MASTERED. As was the case in Study II (see Fig. 7) there seemed to be no important difference between the Low and High Raven groups.

Fig. 19 contains TRIALS TO MASTERY as a function of PASSAGES MASTERED. As noted earlier the first test was counted as a trial so the smallest possible value for the TRIALS TO MASTERY variable was 1. In general these data replicate the data for Study II (see Fig. 8) except the Low Raven group in Study III tended to require more trials to reach mastery. It may be remembered that the Low Raven group tended to go faster on the first trial (from Fig. 17) and this may account for why they required more trials on the average than the High Raven group.

Fig. 20 contains TOTAL TIME as a function of PASSAGES MASTERED. The data from the Low Raven group is a almost perfectly coincident with the data from the

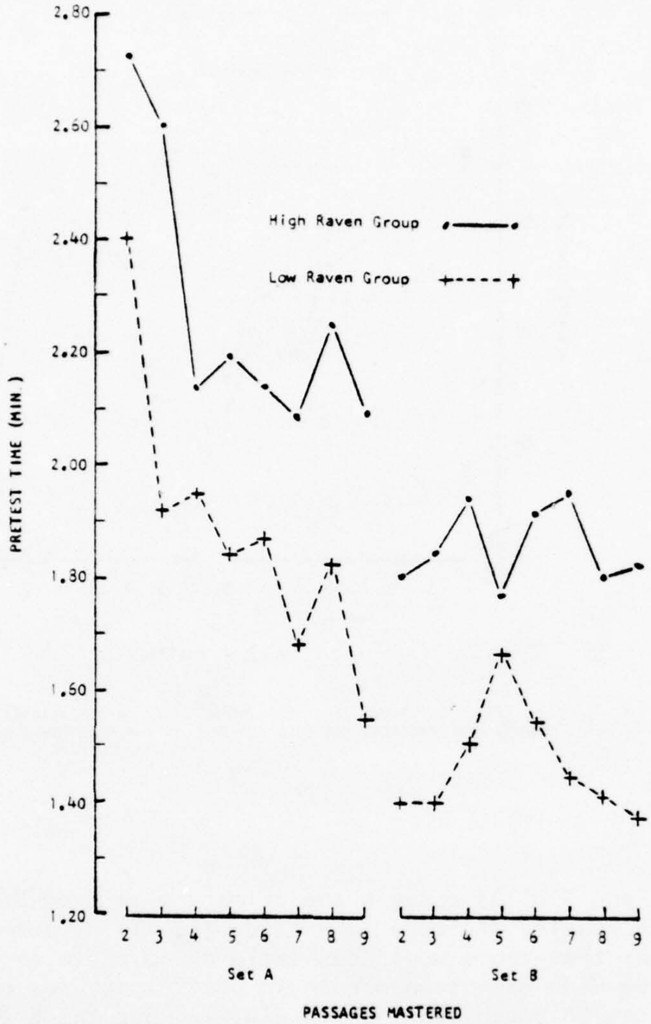


Fig. 17. Study III. PRETEST TIME, in min., as a function of PASSAGES MASTERED for the Low and High Raven Groups.

High Raven group. These two groups required about 1265 min. (21.1 hrs.), on the average, to finish all 160 passages. Compared to the two groups in Study II, one group of those two required 8% less time and the other group required 16% less time than the Study III groups. The data in Fig. 20 indicate again that the Low Raven group tended to master the passages at the same rate as the High Raven group.

Fig. 21 contains POSTTEST HIGHER RGR as a function of PASSAGES MASTERED. Again, these data tend to replicate the corresponding data in Study II (see Fig. 10). There was little or no difference between the Low and High Raven groups. Again, these two groups in Study III tend to start and finish slightly lower than the two groups in Study II.

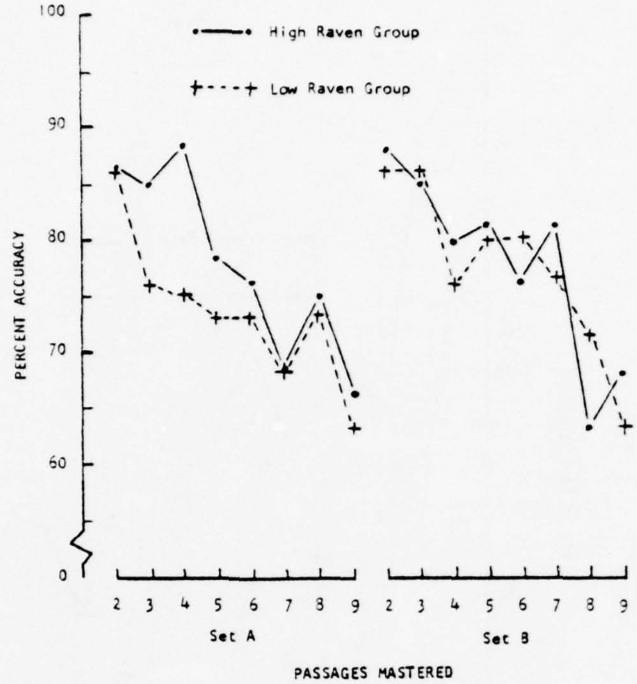


Fig. 18. Study III. PERCENT ACCURACY on the pretest as a function of PASSAGES MASTERED for the Low and High Raven groups.

Fig. 22 and Fig. 23 contain the primary data from Study III. These data replicate the results of Study I and Study II. To reiterate, the research hypothesis was that there would be little or no difference between the Low Ravens and the High Ravens when gain in proficiency was measured with the number of passages held constant, as in Fig. 22, but there would be a difference in favor of the High Raven group when gain in proficiency was measured with total time held constant, as in Fig. 23. The research hypothesis again received no support. There was very little difference between the two groups in Fig. 22 or Fig. 23 but what difference there was tended to favor the Low Raven group as was also the case in Study II.

As was the case in Study II, the data in Fig. 24 will be presented to determine whether the lack of support for the research hypothesis also held for another indicant of proficiency, POSTTEST HIGHER RGR. The data in Fig. 21 and Fig. 24 contain data which are parallel to the data in Fig. 22 and Fig. 23 except the indicant of proficiency is POSTTEST HIGHER RGR instead of PRETEST RGR. The POSTTEST HIGHER RGR Data in Fig. 21 and Fig. 24 also do not support the research hypothesis. The data in Fig. 24 tend to replicate the corresponding results in Study II (see Fig. 13) and do not indicate superiority for the High Raven Group; instead the data again indicated a slight superiority for the Low Raven group.

As was also done in Study II, the individual POSTTEST HIGHER RGR data has been presented for the practice passage that was administered at the beginning

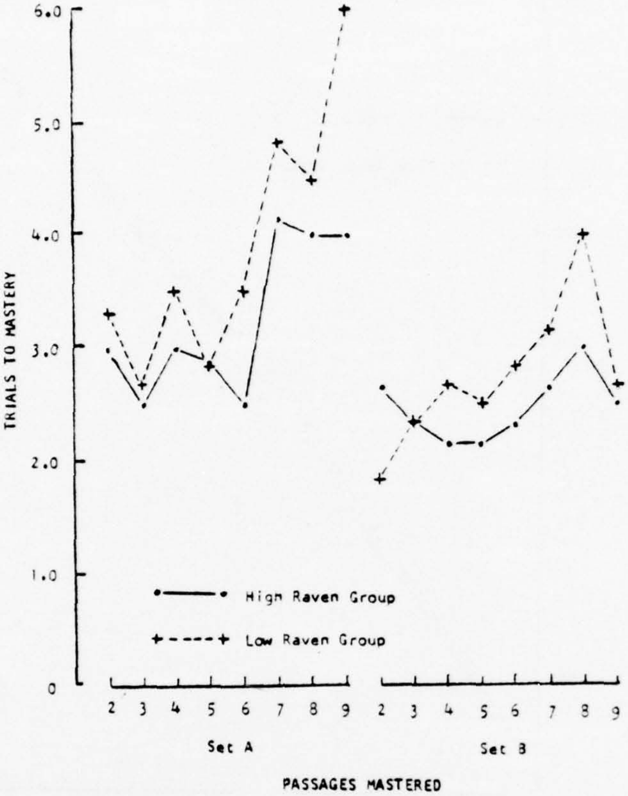


Fig. 19. Study III. TRIALS TO MASTERY as a function of PASSAGES MASTERED for the Low and High Raven groups.

of each training session. Fig. 25 contains these data for each session. The largest difference between two individuals is for S2 and S6. Also included in Fig. 25 are the data for another two graduate students, different from the two in Study II. These latter two students followed the same procedures as did the corresponding two graduate students described in Study II. Again, these two graduate students gained much more in a shorter period of time than the 6 high school students. The drastic differences in gain (note the slopes of the curves) between individuals, with respect to gain in POSTTEST HIGHER RGR, tend to disappear again, however, when they are related to PASSAGES MASTERED--presented in Fig. 26. These data replicate the data in Study II (see Fig. 14 and Fig. 15). The gains of the fastest graduate student in Fig. 26 are approximately equal to the gains of the high school students. Again, this was the type of result expected but not obtained for the PRETEST RGR data when comparing the Low and High Raven groups. It was expected that: (a) the gains would be different under equal time conditions but the same for equal mastery conditions, and (b) that the fastest gainers under the equal time conditions would be the students in the High Raven group.

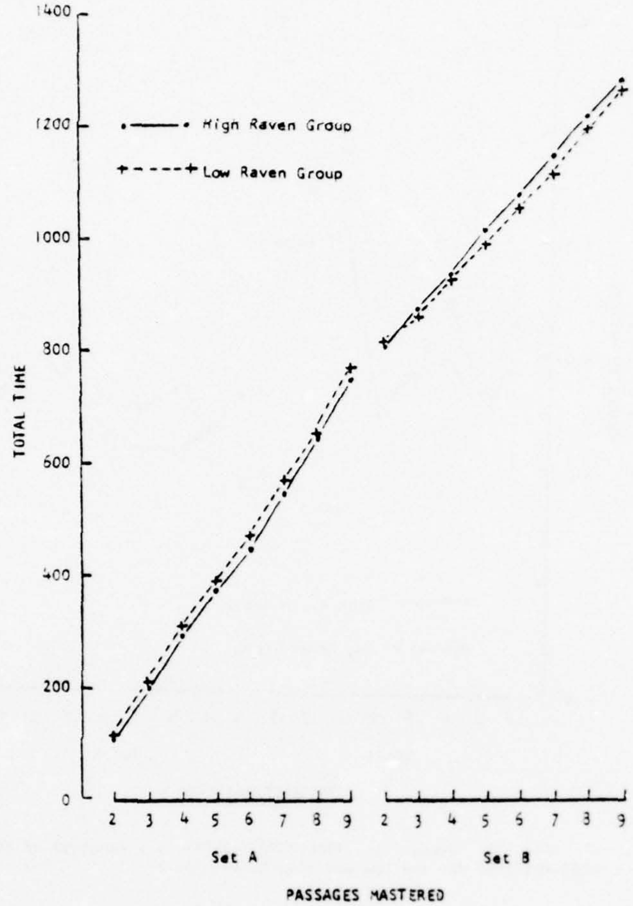


Fig. 20. Study III. TOTAL TIME spent working on the passages as a function of PASSAGES MASTERED for the Low and High Raven groups.

RAVEN RESULTS AND DISCUSSION

As was the case in Study II, the Raven was again administered toward the end of Study III using the same procedures as were used in Study II. To reiterate, the students were paid 10¢ as soon as they answered an item correctly. If they answered the item incorrectly, they were forced to wait one min. before giving their second answer; they were paid 5¢ if they answered the item correctly the second time.

These Study III students had taken the Raven twice in September along with the Study II students. Then, they took it again in December. Finally, they took it again in April under the pay conditions. Thus, they received five different scores on the Raven. The first score was a regular Raven score. The second score came from the second administration of the test where they knew which items they answered wrong the first time; but they did not know which alternative they chose or which alternative was the wrong one. The third

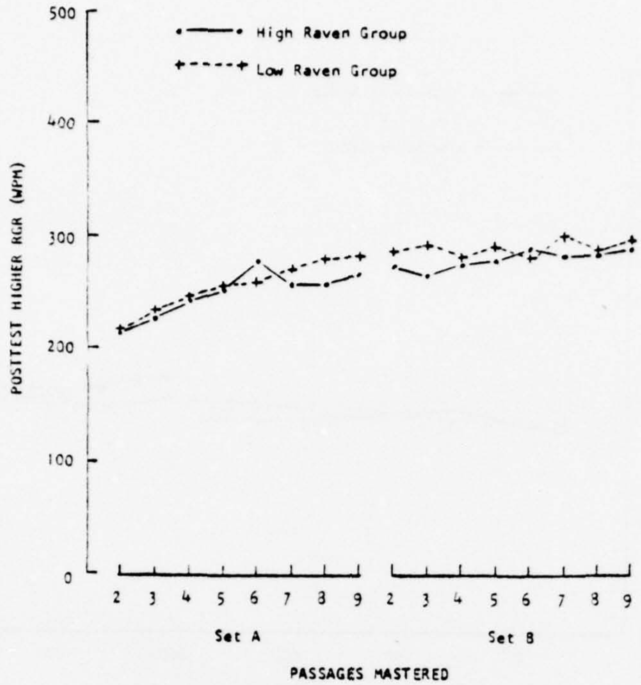


Fig. 21. Study III. POSTTEST HIGHER RGR, in Wpm, as a function of PASSAGES MASTERED for the Low and High Raven groups.

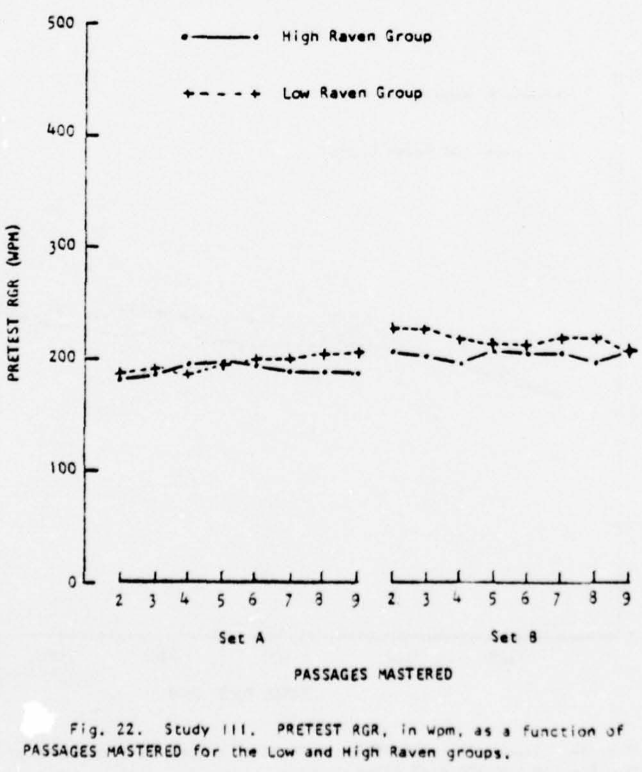


Fig. 22. Study III. PRETEST RGR, in Wpm, as a function of PASSAGES MASTERED for the Low and High Raven groups.

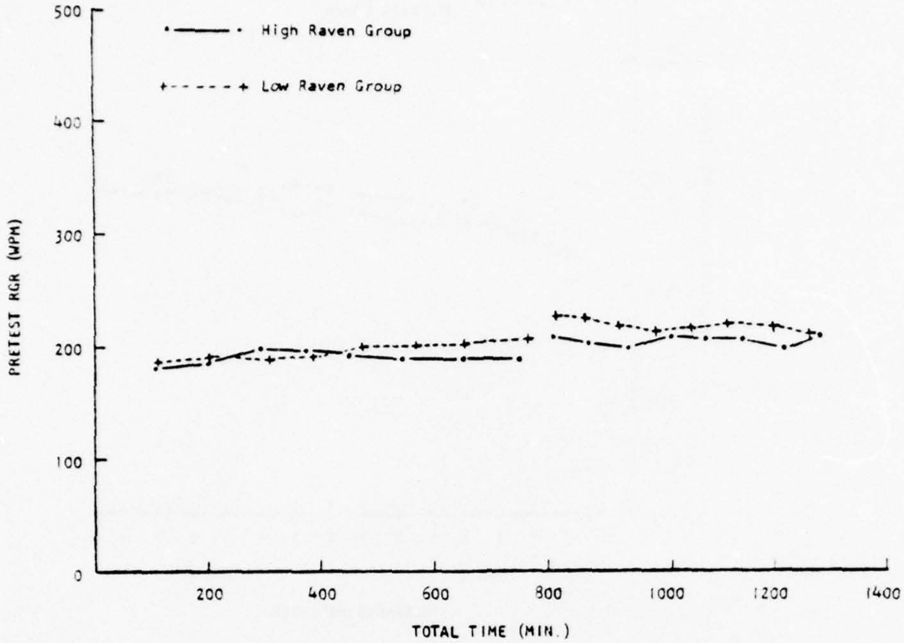


Fig. 23. Study III. PRETEST RGR, in Wpm, as a function of TOTAL TIME, in min., for the Low and High Raven groups.

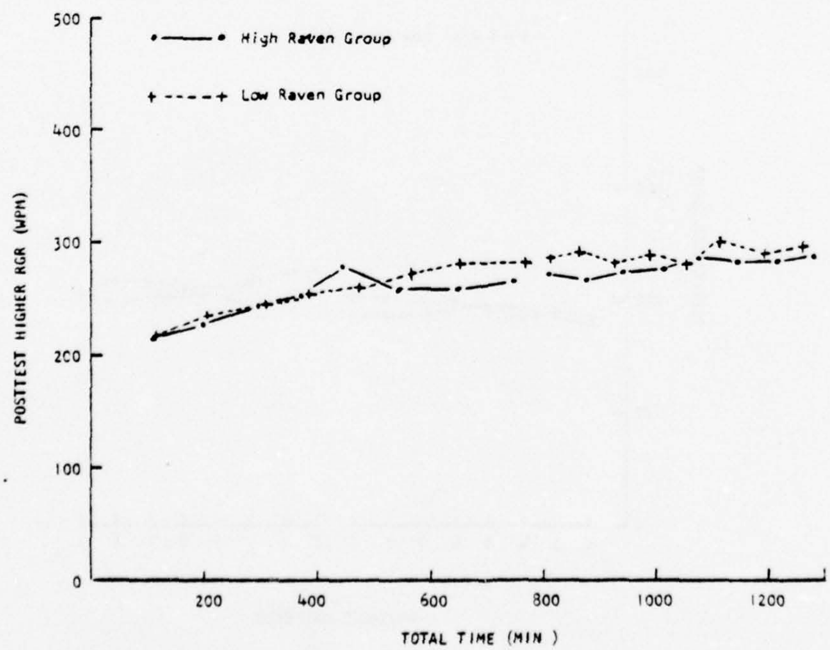


Fig. 24. Study III. POSTTEST HIGHER RGR, in Wpm, as a function of TOTAL TIME, in min., for the Low and High Raven groups.

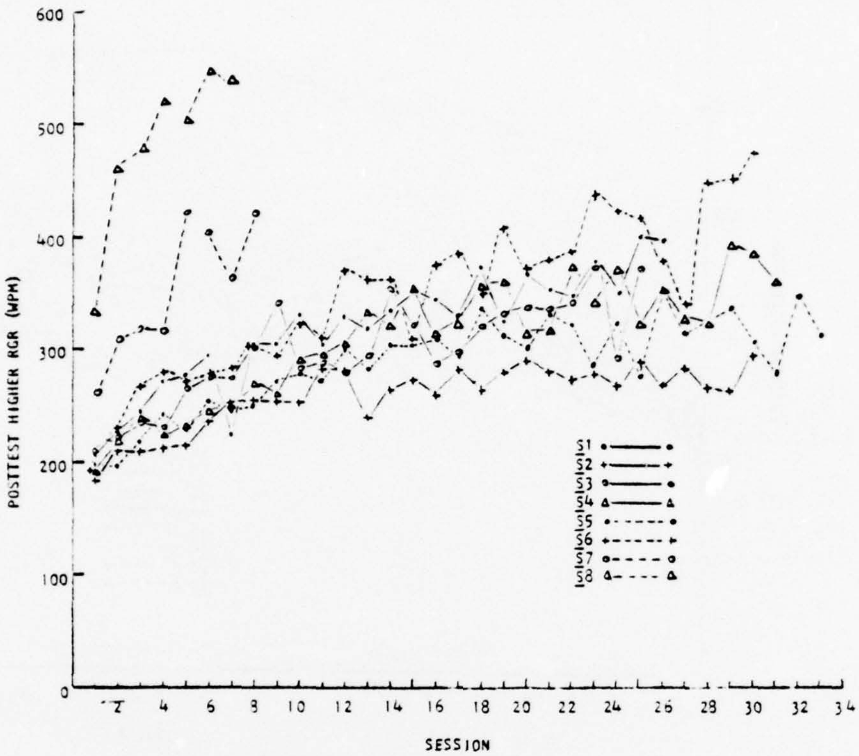


Fig. 25. Study III. POSTTEST HIGHER RGR, in Wpm for the practice passage as a function of the training session attended for each subject.

score came from the third administration where they knew which items (but not which alternative) they answered wrong the second time. The fourth score came from the fourth administration where they were given no feedback from previous administrations but they were paid 10¢ for each correct answer. The fifth score also came from the fourth administration but an incorrect item was changed to a correct item for the fifth score as long as the item was answered correctly the second time. The results for each individual on each scoring are presented in Table 3. The scores are grade level equivalents,as described earlier.

The 6 students were selected after the second scoring when the High Ravens (S1, S2 and S3) ranged from 7 to 9 and the Low Ravens (S4, S5, and S6) ranged from 4 to 6. After the fifth scoring, it may be noted that the High Ravens ranged from 11 to 12 and the Low Ravens ranged from 9 to 12. The difference between the means of the High and Low Raven groups on the second scoring was 3.6; by the fifth scoring this difference was only 1.0. Thus, after motivation supposedly had been maximized and after impulsivitity supposedly had been minimized, then the differences between the two groups became much smaller, being of the order of only one grade level difference.

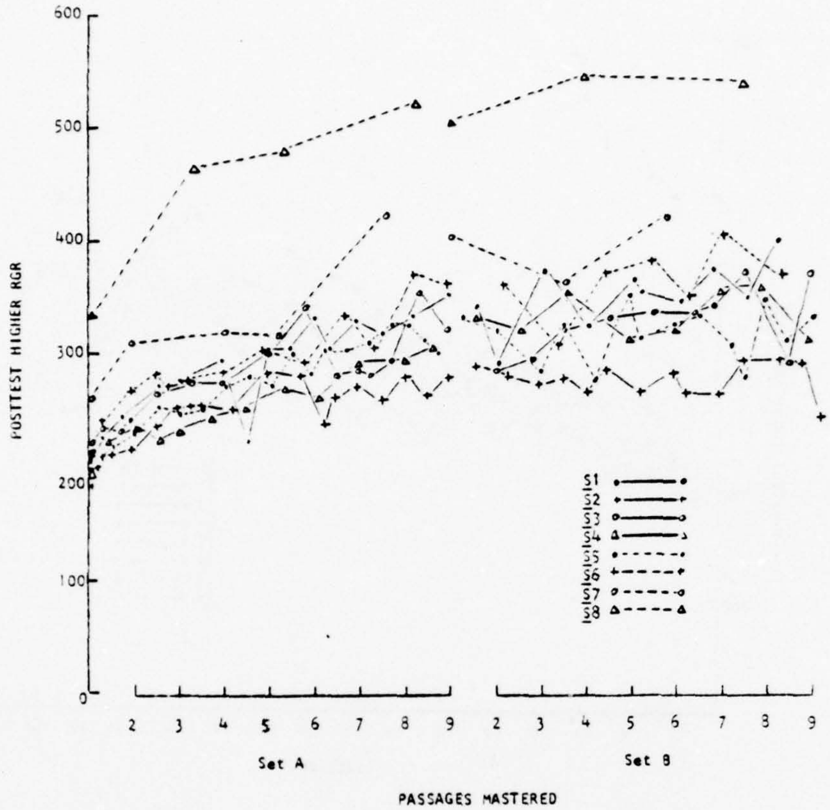


Fig. 26. Study III. POSTTEST HIGHER RGR, in Wpm, for the practice passage as a function of PASSAGES MASTERED for each subject.

Table 3

Raven Scores in Grade Equivalents

for the 6 students in Study III

	Scoring Occasion				
S	First	Second	Third	Fourth	Fifth
1	6	7.	10	9	11
2	9	9	10	11	12
3	5	9	9	9	11
4	2	6	7	10	12
5	5	4	6	8	10
6	3	4	6	7	9

TRAINING EFFECTIVENESS

INTRODUCTION

The primary purpose of the research reported under Studies I, II, and III, was to compare the gain in reading ability of the Low Ravens with the High Ravens. Thus, the research design focused upon this comparison. A secondary concern, however, was the effectiveness of the training procedure itself as a method for improving reading ability. If the training procedures did not in fact increase reading ability, this would have a crucial effect upon the interpretation of the research results. This section will focus upon the supplementary data that was collected relevant to the effectiveness of the training technique as a facilitator of gain in general reading ability.

The data from the practice passage in Studies II and III leave no doubt but that the individuals in fact learn to read that passage more efficiently. However, this type of task is far removed from ordinary reading since it involves primarily increasing the speed at which the correct word in a sentence can be recognized when the sentence is part of a passage that has been read so many times that it is probably almost memorized. Compared to the POSTTEST HIGHER RGR data for the practice passage, the POSTTEST HIGHER RGR data for the 160 passages is closer to ordinary reading. Each passage is not mastered more than once, as was the case for the practice passage. Yet, this task is also far removed from ordinary reading. It involves the repeated reading of the same passage until all items are answered correctly, i.e., the passage is mastered. Thus, just because there is gain on this variable, there is little cause to infer that this reflects gain in general reading ability. The best data as discussed earlier, for inferring gain in general reading ability comes from the first trial RGR variable, i.e., the PRETEST RGR data. These data reflect how well the student reads a passage that the student has never seen, read, or heard before. This task is much more similar to what happens during ordinary reading. Thus, these data, from Study II, were analyzed in more detail.

Another way used to evaluate the effectiveness of the training upon general reading ability, was to administer standardized achievement tests. This was an evaluation of the degree to which the gain in skill on the computer administered reading tasks transfered to other reading situations.

The results from the aforementioned Study II data and transfer data will be presented in the two subsequent sections.

STUDY II DATA

The PRETEST RGR data for each of the subjects in Study II are presented in Fig. 27. These data were obtained by calculating: (a) the mean of the two medians at each grade level for each group of 5 passages (this step was the same initial step as was described earlier when the Low and High Raven analyses were described), and (b) the mean of the grade 2 and 3 means was calculated (called Block 2.5), the mean of the grade 4, 5, and 6 means was calculated (called Block 5.0) and finally the mean of the grade 7, 8, and 9 means was calculated (called Block 8.0). This procedure was followed for both the Set A and Set B passages for each student.

It may be noted in Fig. 27 that for each Block, indicating equal passage difficulty, each person showed a gain between the Set A and Set B passage. These data suggest that each of these six students gained in their ability to read the

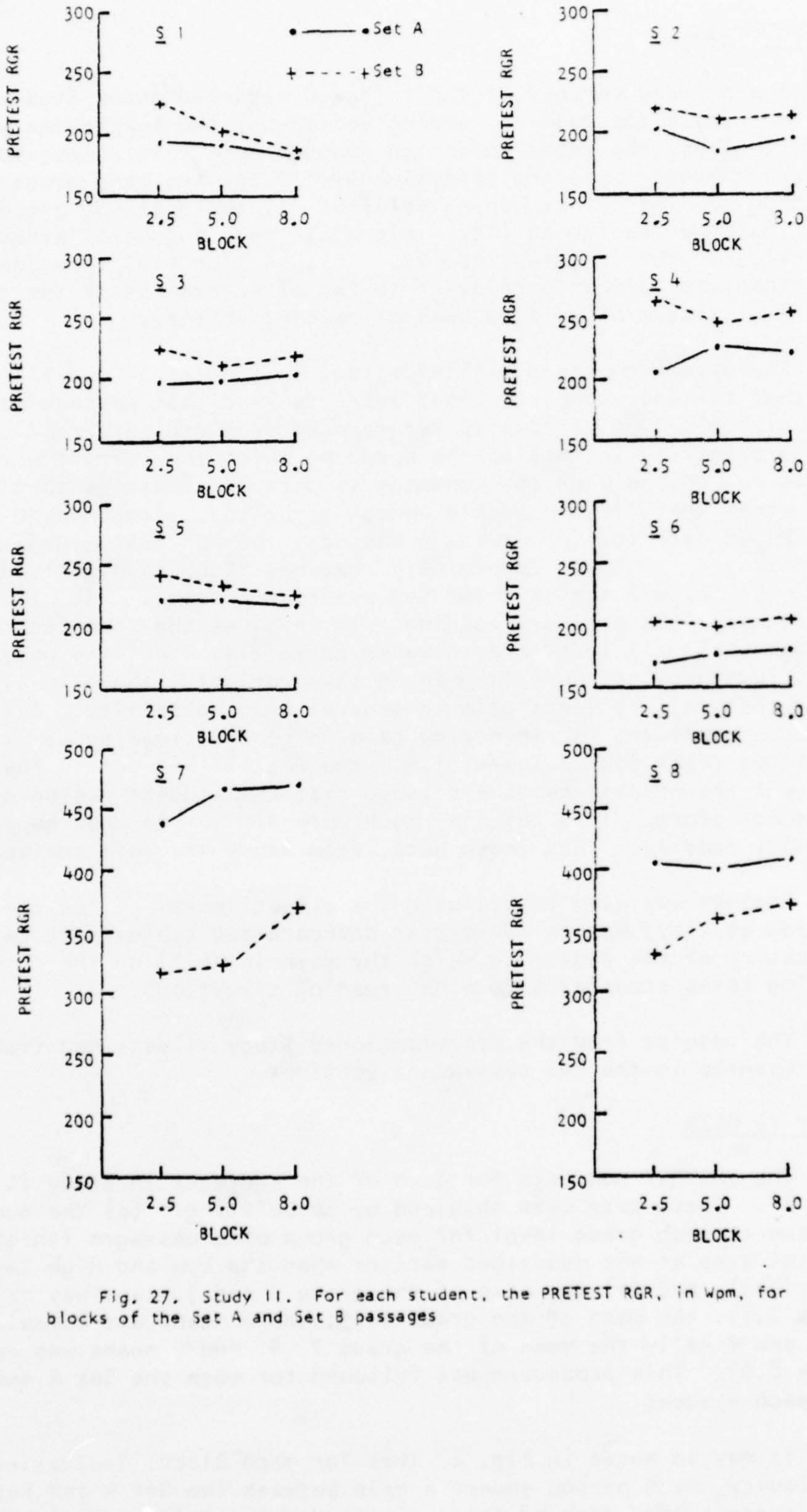


Fig. 27. Study II. For each student, the PRETEST RGR, in Wpm, for blocks of the Set A and Set B passages

sentences in the passages.

The gains in Fig. 27 may have been caused by the reading practice training. But they could also be trivial gains. Furthermore, there could be several alternative hypotheses that could just as easily explain these data. The first alternative hypothesis is that the gain in skill was specific to the task or the computer equipment used in the research. Evidence relevant to this hypothesis and relevant to whether the gain is trivial or not comes from the NRS test. This test was administered prior to the training, as described earlier, and also subsequent to the training, i.e., at the end of the training sessions. Table 4 contains the median grade equivalent scores for the pre and post administrations of the NRS test, together with the gain scores. The gain for the 6 students was 3.6 grade equivalents,

Table 4

Pretest to Posttest Gains in Study II

	Pretest	Posttest	Gain
NRS Test (Grade Equivalents)	4.5	8.1	3.6
NRS Test (Estimated Reading Rate in words per min.)	175.5	205.5	30.0
PRETEST RGR (set A, Block 2.5 for Pre and Set B, Grade Block 8.0 for Post)	201.5	216.5	15.0
PRETEST RGR (Median of first 5 passages for Pre and Median of last 5 passages for Post)	195.0	215.5	20.5

starting with 4.5 on the pretest and going to 8.1 on the posttest. Part of this gain is no doubt due to repeated administrations of the test. Part of the gain also could be due to regression effects but this is not likely since the NRS test is a sequential type of test where extremely high or extremely low scores indicate a higher or lower forms of the test should be given. The 6 subjects made extremely low scores on the NRS 3A test and were subsequently administered the NRS 2A test. They made neither extremely high or extremely low scores on this test. Therefore, regression to the mean should not be a major factor contributing to this gain. Since there were not enough students who made equal reading scores and still differed widely on the Raven to form a control group, it is difficult to assess the amount of gain due to repeated testing.

There is a replicative type of evidence for the size of the gain as indicated by the NRS test. Using the manual for the test, the NRS grade equivalent scores can be converted into rate of reading scores in words per min. This rate is supposed to be the rate at which an individual would ordinarily read a passage that was relatively easy for the individual to read. These data are also presented in Table 4. Notice that the median pretest rate was 175.5 words per min. and the median posttest rate was 205.5 words per min., reflecting an absolute gain of 30 words per min or a 17% gain. The above gain can also be compared to the gain in the RGR variable, discussed earlier. The equation for calculating the RGR, Equation 1, was developed using the same formulas as was the NRS rate of reading scores mentioned above. Therefore, the RGR scores should be comparable to these rate of reading scores. The only major difference is that the NRS test requires that a pencil be used to mark the box in front of the correct items while the computer task from whence the RGR scores came, requires a key on the left or on the right to be depressed to indicate choices.

The mean of the PRETEST RGR scores on Block 2.5 for Set. A, in Fig. 27, may be used as an indicator of pre training ability and the mean of the Block 8.0 for Set B in Fig. 27 may be used as an indicator of post training ability. The medians of these individual means are also presented in Table 3. The pre-training median was 201.5 words per min and the posttraining median was 216.5 wpm for a gain of 15 wpm or a 7% gain. This 15 wpm gain is not far from the 30 wpm gain on the NRS. Another way of analyzing these same data would be to let the median RGR of the first 5 passages reflect pretraining ability and the median RGR of the last 5 passages reflect post training ability. This analysis is closer to the beginning and the end of training, and may reflect the gain better. These data are also included in Table 4. The median of these 6 pretraining indicators was 195.0 wpm and the median of the post training indicators was 215.5 wpm giving a gain of 20.5 wpm for a 11% gain. These gain data are close to the NRS data, i.e., the 30 wpm or 17% gain.

The gains using the pretest RGR data and the gains using the NRS seem to be relatively comparable. Thus, it appears reasonable to suggest that this reading skill training produced about a 10-15% gain in reading efficiency for these individuals. This increase may also be interpreted as a gain of around 3-4 grade equivalents on a standardized test. However, inferrences about the size of the gain should be interpreted cautiously since there was no control group. Also, it should be pointed out that the PRETEST RGR data from the two graduate students also showed gains; these data are also presented in Fig. 27. Since these two graduate students showed large gains on the first trial RGR data as a result of training, should we also conclude that they became better readers as a result of this brief training? Their gains were greater than the high school students. It does not seem reasonable to expect that these two graduate students became better readers simply by completing this reading training practice on these grade 2 through grade 9 passages. Their gain seems to be more reasonably restricted to this task and not transferable to their regular reading. If this is in fact the case, then is it reasonable to expect the gains of the 6 high school students to transfer to their regular reading? It could be that all of the reading progress is specific to this particular type of task, and does not transfer to regular reading.

In order to further assess the effectiveness of the training procedure another data collection effort, described in the next section, was initiated.

TRANSFER DATA

Introduction. Additional testing was conducted in order to get an indication of whether the gain in reading performance experienced by the high school students was a gain in general reading ability or not. Since the NRS test provides a task that is almost exactly the same as the training task itself, it is quite possible that the gain in general reading ability indicated by the NRS test is artifactual. It could be that this gain is quite specific to the task and does not transfer to ordinary reading or to other reading tasks. Thus, the following data collection effort was designed to provide an answer to this question regarding general or specific gains in reading ability.

The full set of the Gates-MacGinitie Reading Test was administered to all of the Study II and Study III Ss. This test has three subtests--Speed and Accuracy, Vocabulary, and Comprehension--which provided data relevant to the general gains, or the degree to which the skills learned by the high school students transfer to other reading situations.

Data were also collected relevant to reading rate. It could be that these students became more accomplished readers only in the sense of being able to read faster. That is, they may not be able to comprehend any more than they could prior to training but they may be able to comprehend the same material in less time. They may have improved the efficiency with which they transform the words in sentences into the complete thoughts the words represent (see Carver, 1978). Thus, it seemed important to collect data regarding typical, normal, or ordinary reading rate; a set of Reading Rate measures was developed, as described later.

Alternate forms of the Reading Rate measures and the Gates-MacGinitie tests were administered during January and April. For the Study II Ss, the January testing took place soon after they had finished training, POSTTESTING, and the April testing took place long after they had finished training, DELAYED-POSTTESTING. For the Study III Ss, the January testing was at the beginning, PRETESTING, and the April testing was at the end of the training, POSTTESTING.

The data to be presented in this section provided better evidence relevant to whether the gains reflected by the reading efficiency measures and the NRS tests were replicable by other measures.

Test Procedures. The six Study II students were paid $6.00 each to return for 2 hours of testing in January. At the outset, four reading rate measures were administered. Four 100-word passages, two grade 1 and two grade 4, were given to the student to read. The student was asked to read the passage at his or her normal rate and was told that no test would be given because the experimenter was only interested in measuring normal reading rate. The experimenter measured the elaspsed time to read each passage using a stop watch. Altogether, there were 4 passages at grade 1 difficulty and 4 at grade 4 difficulty (these same passages were also used in a previous experiment, Carver, 1977b); difficulty level was measured using the Rauding Scale. Two of the passages at each grade were administered during the subsequent April testing. The order of administration of the passages and the specific passages administered to the group were counterbalanced over the pre and post testing using an incomplete Latin-square.

After the reading rate measures had been administered, the Speed and Accuracy Subtest, the Vocabulary Subtest, and the Comprehension Subtest of Survey D. Form 1 of the Gate's MacGinitie Reading test were administered.

The six Study III students were administered the same tests as described above for the six Study II students. The four reading rate measures and the Speed and Accuracy subtest and the Vocabulary subtest of the Gates-MacGinitie were administered at the beginning of Session 3 and the Comprehension Subtest of the Gates-Mac Ginitie was administered at the beginning of Session 4.

The same measures that were administered during January were also administered again in April using the same procedures. However, for the April testing, Survey D, Form 2 of the Gates MacGinitie was used, and each student was asked to read four different passages for the Reading Rate measures, as described earlier.

As described earlier, the NRS testing, Form A, was administered in September to both the Study II and Study III Ss. The NRS Form B testing was administered in December to both groups. The NRS Form A testing was also administered again during the April Testing.

An overview of the testing and the training schedule for both the Study II and Study III students is presented in Table 5. Since the Study II students were given the NRS, the Reading Rate measures, and the Gates-MacGinities again in April, long after they had finished their training in December, this testing was referred to as the DELAYED-POSTTESTING for them. Since the Study III students were given the NRS test in September long before they began training in January,

Table 5

An Overview of the Testing and Training

Schedule for the Study II and Study III Students

Month	Study II Students	Study III Students
September	PRETESTING: NRS	ADVANCED-PRETESTING: NRS
October	(TRAINING)	
November	(TRAINING)	
December	POSTTESTING: NRS	PRETESTING: NRS
January	POSTTESTING: Rate, Gates	PRETESTING: Rate, Gates
February		(TRAINING)
March		(TRAINING)
April	DELAYED-POSTTESTING: NRS, RATE, GATES	POSTTESTING: NRS, RATE, GATES

this testing was referred to as the ADVANCED PRETESTING. Although the NRS was given in December and the Rate and Gates measures were given in January, they seemed to be close enough in time to lump them together as POSTTESTING for the Study II students and PRETESTING for the Study III students.

Results and Discussion. Fig. 28 contains the results for the NRS test. Each data point represents the median of the six grade level ability scores on the NRS test. The pre to postest gain for the Study II students, also noted earlier, was from 4.5 to 8.1, i.e., a gain of 3.6 years or grades. The corresponding gain for the Study III students was from 4.7 to 8.6, i.e., a gain of 3.9 years or grades. It may be noted that the Study III students almost perfectly replicated the pre to posttest gain of the Study II students. There was a small gain for the Study III students from the ADVANCED PRETESTING to the PRETESTING, 1.3 which suggests some improvement due to repeated testing. There was no change for the Study II students from the POSTTESTING to the DELAYED-POSTTESTING, which suggests stability for the gains over time.

These data seem to reinforce the conclusion that there was a gain in

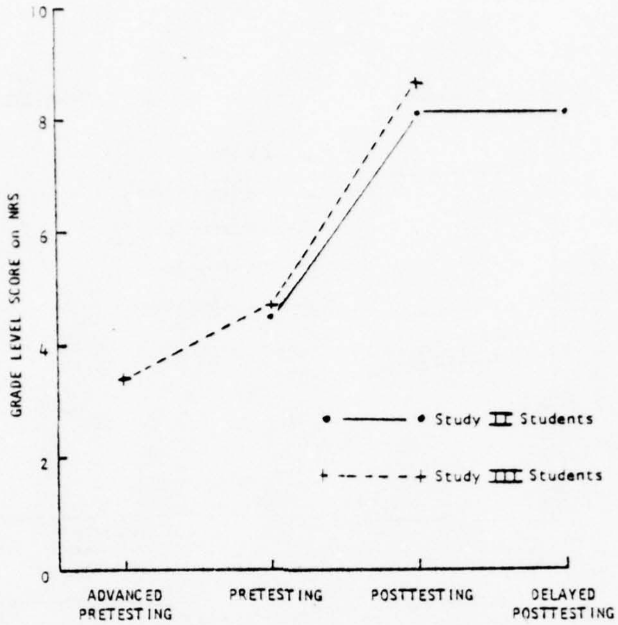


Fig. 28. Grade level scores on the NRS for the Study II and Study III students when tested under various pre- and post-training conditions.

reading performance as a result of training at least as measured by a test that involved almost exactly the same type of reading task as was involved in the training itself.

Fig. 29 contains the Gates-MacGinitie test scores as well as the NRS test scores for comparison purposes. The left side of the figure contains the results for the Study II students, and the right side contains the results for the Study III students. Each side contains the results for the December/January testing versus the April Testing, i.e., the same testing periods.

For the Study II students, the December/January testing is the POSTTESTING and the April testing is the DELAYED-POSTTESTING. The scores for the Study II S should be stable between the two testing, and they were. The POSTTESTING scores on all the Gates subtests are closely clustered around grade level 5 while the corresponding score for the NRS is about 8. Thus, the high gain to about grade level 8 reflected by the NRS for the Study II students in Fig. 28 does not seem to generalize to other indicators of reading ability. It is not surprising that the Gates Vocabulary scores were not higher, i.e., around grade 8, because the training was not focused upon vocabulary training. However, if the students had learned to read more efficiently it would seem that this increase in ability would have resulted in approximately grade 8 scores on the Speed and Accuracy subtest or even the Comprehension subtest since it is also a speed test. (Note: The Gates-MacGinitie test manual gives two grade equivalent scores for the Speed and Accuracy subtest, the Speed score

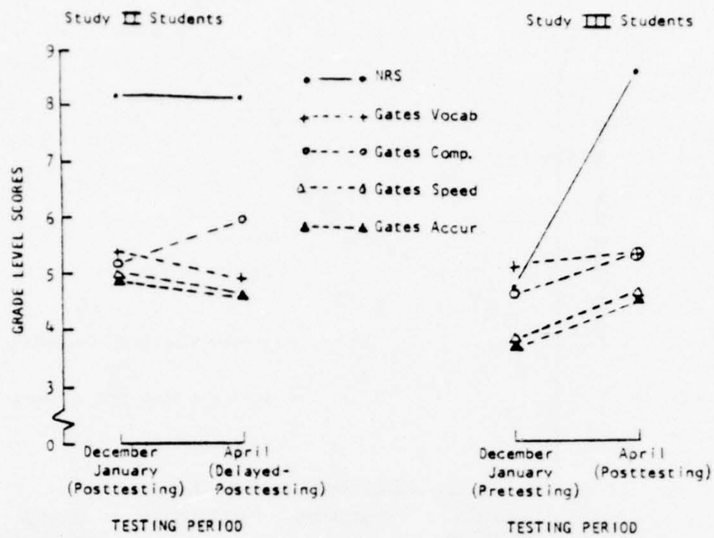


Fig. 29. Grade level scores on the NRS reading test and the subtests of Gates-MacGinitie reading test for two testing periods; one graph is for the Study II students and the other graph is for the Study III students.

is based upon the number attempted and the Accuracy score is based upon the number correct.) The data from the Study II students do not support the idea that the training produced a gain in general reading ability. On the contrary, these data suggest that the training used does not transfer to other reading tasks. The April testing, i.e., the DELAYED-POSTTESTING data simply replicated the POSTTESTING data thus suggesting that these results were reliable; there was no training administered between the December/January testing and the April testing for the Study II students, and there was also very little change in the scores during this period.

For the Study III students, the December/January testing was the PRETESTING data and the April testing was the POSTTESTING data. These data indicate that all the measures of reading ability are in agreement in estimating the reading ability of these students to be around grade 4 or 5 prior to training, i.e., the PRETESTING data. After training however, the NRS shows a drastic gain to about grade 8 while the Gates-MacGinitie subtests show little or no gain with all scores clustering again around grades 4 and 5.

The preceding transfer data may be summarized as follows:

1. The subtests on the Gates-MacGinitie reading test indicated that both groups of students had about a grade level 5 average in reading ability prior to and also subsequent to training, i.e., the Gates-MacGinitie indicated no gain from grade 5 ability as a result of the reading training.

2. The NRS test results replicated the Gates-MacGinitie test results prior to training i.e., both groups of students average around grade level 5 ability prior to training, but the NRS test results

do not replicate the Gates-MacGinitie subsequent to training since the NRS scores showed a gain to around grade 8 for both groups.

3. These data may be interpreted as indicating that the reading training was not effective in increasing general reading ability but was only effective in increasing the ability to perform the programmed prose type of reading task that was used in training and employed in the NRS test.

The reading rate data for the first, second, third, and fourth passages were analyzed separately, initially. That is, the median rate for the six students for each of the four orders was calculated and then the mean of these four median rates was calculated as a measure of the typical rate under each condition. For the Study III students, the resulting value was 121 Words per minute for the PRETESTING and 144 Words per minute for the POSTTESTING. For the Study II students, the resulting value was 154 Words per min. for the POSTTESTING and 168 Words per min. for the DELAYED-POSTTESTING. Thus, the training effect for the Study III students seemed to be a gain of 34 Words per min. to bring them from 121 to 155 Wpm and make them equal to where the Study II students were when they finished their training, i.e., 154 Words per min. Yet, this gain is difficult to interpret since there was no direct control. The Study II students gained 24 Words per min. with no training condition in between and this detracts from directly interpretating the rate gain of the Study III students as being attributable to the training.

The reading rate data are not clear-cut with respect to providing evidence either for or against the training being effective in producing a gain in reading ability that transfers to typical reading situation.

SUMMARY OF TRAINING EFFECTIVENESS RESULTS

Although the reading rate data may leave some room for doubt, the results of the Gates-MacGinitie test indicate that the reading training used in this research was not effective in producing an increase in general reading ability. If there was such a gain, it would seem to have been small at least in comparison to the large gain that was artifactually reflected by the NRS test. Based upon all of the evidence presented, it seemed reasonable to conclude that the 60-70 hours of reading training administered to each student by the PLATO computer terminal was minimally effective in producing gains in general reading ability. There were in fact large gains in the ability to perform the programmed prose type of reading task that was utilized in the training. However, there was evidence that indicated that these increases in ability were largely specific to this task.

SUMMARY AND CONCLUSIONS

The data presented from Study I, and Study II, and Study III all indicated no support for the hypothesis that poor readers consist of two types--those that read poorly because of meager reasoning ability and those that read poorly because of insufficient reading practice. From these data it may be concluded that if a group of students who read equally poorly are divided on the basis of their reasoning ability, the low group will gain just as much from reading training as the high reasoning ability group.

Superficially, the data were quite definitive with respect to the preceeding conclusion. Data supplementary to that outlined above, however, renders this conclusion questionable at best. Part of this supplementary data comes from the Raven Progressive Matrices Test. It was used as a screening instrument for separating students into low and high intellectual ability. The reliability and validity of the data from this test was questioned early in the research and additional data was collected. The first data collected provided evidence for the validity of the Raven; the Verbal and Performance scores on the WISC test provided constructive replication data (see Lykken, 1968) for the Raven. The Low Raven group tended to score low on both the Verbal and Performance sections of the WISC thus suggesting low potential with respect to gain in reading ability as a result of training. The High Raven group tended to score higher on the Performance section as compared to the Verbal Section thus suggesting high potential with respect to gain in reading ability as a result of training. However, the validity of the Raven data was put in jeopardy from another direction. The test was repeatedly given under several conditions wherein possible variation in motivation and impusivity was designed to be reduced to a minimum. Under these conditions, the differences between the Low Raven group and the High Raven group almost evaporated completely. It was also suggested that the WISC data would be similarly affected by this type of testing control. Thus, it is reasonable to question whether the difference between the two reasoning ability groups was large enough to be important. If the differences between the Low Raven groups and the High Raven groups were in fact very small with respect to reasoning ability, then the original hypothesis guiding this research was not tested and the preceeding conclusion is questionable at best.

Another qualification that detracts from the validity of the conclusion drawn at the outset of this section, involves the training effectiveness results. If the training was of questionable effectiveness in producing a general gain in reading ability, then any conclusions drawn from the research results about the relationship between reasoning ability and gain in reading ability also become highly questionable from this standpoint.

There was one aspect of the research results that indirectly supported the theoretical rationale given at the outset. Gain in the efficiency with which the students could perform the programmed prose task on the practice passage showed great variability between students when gains in performance were considered with respect to the number of training sessions or the total amount of time engaged in the reading training. However, when the gains in performance were considered with respect to the number of reading passages that had been mastered to a quality criterion, then the individual differences in gain vanished.

From this research data in its entireity, it seems reasonable to conclude the following:

1. The Raven Progressive Matrices Test is of questionable validity for

measuring the intellectual ability of poor reading, high school students unless it is given under special conditions wherein motiviation can be manipulated to a high level and impusivity can be reduced to a low level.

2. Use of the programmed prose reading task as a training procedure for poor reading, high school students does in fact produce gains in ability to perform this type of reading task. Unfortunately, the gains in this skill, or ability, seemingly do not transfer, to any substancial degree, to other types of reading situations that may be more common to typical reading situations.

3. Individual differences seem to approach a vanishing point with respect to gain in the ability to perform the programmed prose reading task when gain is not related to time spent mastering the task but is related to the number of successfully mastered tasks.

IMPLICATIONS AND RECOMMENDATIONS

This section contains speculations that in some areas go well beyond the data collected in the present research.

Programmed prose, as a training technique for improving reading ability, seems to have limited potential. The present research indicates that it is not a panacea for improving the reading ability of poor reading, high school students. The fact that these students showed gains on this reading task indicates that they are capable of learning but that the skill learned does not readily transfer to other reading tasks. Thus, it would seem that if these students are capable of learning, it is still the task of education to figure out how or what they need to be taught in order for them to become average readers instead of poor readers. The present research results suggests that simple practice in reading words, i.e., determining which words belong in sentences, is not a sufficient condition for improving reading ability in general. Thus, it would seem that reading practice in and of itself is of limited usefulness for these students.

It seems that the general advice given by Smith (1971; 1973) regarding practice and reading improvement needs to be refined. Reading practice under feedback conditions does not seem to have the effect suggested by Smith. Possibly Smith's advice is still good advice for beginning readers, i.e., readers who need decoding practice. However, those readers who have advanced to around grade level 4 or above may not benefit from such practice. It seems more likely that the poor readers who have advanced this far are only going to improve their general reading ability by improving their background knowledge or experience. If reading practice means that the individual is comprehending thoughts never before comprehended, then this type of practice is likely to be effective in increasing general reading abilitv. However. if reading practice simply means that the student is learning which words belong in sentences, by repeated attempts with feedback, this type of practice seems to be of low or zero effectiveness in improving general reading ability. It is possible that certain poor readers have low degrees of reasoning ability and this in turn limits their potential reading ability. However, the fact that the present poor reading,high school students continued to improve on the Raven tests, all the way up to grade 10-12 ability levels, when motivation was increased and impulsivity was decreased, suggests that limited reasoning ability is not likely to be the primary cause of their poor reading. It seems more likely that the cause of poor reading is primarily low background knowledge. Anyone, including college graduates, could be considered as poor readers if given a comprehension test on extremely difficulty material in a subject matter area with which they are unfamiliar.

It would seem that poor reading high school students could be given practice in reading, as was the case in this research, and could be given training in how to think, using some technique approved by Thorndike, and still not become better readers. A more likely possibility for improving their reading ability would be to focus upon educating them i.e., providing direct and vicarious cognitive experiences which provide background knowledge, or schemas, allowing them to better understand what they read or what they are told.

The problem then becomes circular. The best way to become a better reader is to gain more background knowledge about the material that one might be called upon to read, and the best way to gain this knowledge is by reading. The solution to this circularity problem probably lies in subject matter specialization. Rather than trying to get the poor readers to become better readers in general, try to get them to become better readers with respect to a specific subject matter area. For example a high school student who wanted to become an auto mechanic could be given reading instruction using books and manuals dealing with automobile repair. This is the instructional strategy that was advocated and implemented by Sticht (1975) with respect to reading training in the U.S. Army. The present research has not resulted in evidence for a viable alternative to this approach. Probably the best strategy to use to improve the reading ability of poor readers is to concentrate on making these students better readers in a specific subject matter area and give up trying to make them better readers in general. This means that the educational system is likely to be able to help poor reading, high school students become better readers of automobile repair manuals, for example. Yet, no one should expect successful achievement of this goal to result in considerably better general reading ability as reflected by a standardized reading test such as the Gates-MacGinitie.

When the present research was conceived, it was perceived that if the Type II poor reading high school students actually existed and could be isolated using the Raven test, then they could be turned into good readers using modern technology, PLATO IV, to administer reading practice. Thus, it was disappointing from a practical standpoint to find that the reading training resulted in little or no improvement in general reading ability.

The present results may be interpreted as suggesting that the ideas of Thorndike, and the refinements of those ideas as expressed in the Type I and Type II poor reader hypotheses that guided the present research, are either wrong or too simplistic. The idea that simple practice in reading with feedback will redeem poor readers found no support in the present research. Yet, the fact that the poor readers did gain in reading proficiency on the programmed prose task without any direct training in intellectual functioning, suggests that meager intellectual ability is not the sole cause of poor reading as Thorndike seems to suggest. Furthermore, the large gains made by the poor readers on the Raven test, up to grades 10-12 ability, without any direct training in thinking or strategies (other than test taking controls) also seems to provide evidence against the idea that their poor reading was primarily due to meager reasoning ability.

In the beginning, it was noted that Thorndike suggested that if we want better readers then we should develop better ways of teaching people to think rather than concentrating on reading. This approach does not seem to have a solid theoretical rationale supporting it, and empirical evidence to support it seems to be almost exclusively correlational rather than causal or experimental. On the contrary, the present data, although meager, suggest that direct efforts in trying to teach people to think would not likely be a successful way of helping poor readers become better readers. On the other hand, the idea directly investigated in the present research does not seem to hold much potential either, i.e., practicing reading does not seem to improve the reading ability of poor readers. Thus, it seems that the only way left for the vast majority of the poor readers in the United States to improve their reading ability is to improve their background knowledge relevant to what they are reading. Since the best way to increase background knowledge relevant to a specific reading passage is usually to read other material that is relevant to the comprehension of that passage, the potential for helping most poor readers appears to be quite dismal. There appears to be no direct training procedures that will help in any general way since reading improvement can almost be likened unto the difficulty of lifting oneself by one's own bootstraps. Of course, gain in reading ability is not quite that difficult, otherwise there would not be so many good readers. Gain in knowledge and gain in reading ability seem to be highly related in our society, at least by the time the student reaches the point where he/she can comprehend by reading almost everything that can also be comprehended by auding. From this point on, a small gain in knowledge results in a small gain in reading ability and that small gain in reading ability may be used to effect another small gain in knowledge. This theorized interplay between gains in knowledge and gains in reading ability seems to explain why the present research failed to produce a substantial gain in general reading ability. The present research did not focus upon a gain in knowledge by somehow forcing the students to comprehend the thoughts in the reading material. Furthermore, this theorized relationship also predicts failure for training procedures which focus directly upon some type of thinking improvement as a means for increasing reading ability. If gain in reading ability is closely related to gain in knowledge, then general intellectual functioning will have a decidedly secondary influence, if at all.

In closing, the final implications of the empirical results and theoretical ideas presented in this research report will be presented. These implications are quite speculative; they represent the considered opinions of one researcher about an applied area which lacks definitive research results.

The reading training for poor readers who are reading at least as high as the fourth grade level on a standardized reading test should:

1. focus upon increasing subject matter knowledge and increasing reading ability at the same time using the same training material for both purposes.

2. not use training materials that are questionable from a gain in background knowledge standpoint, e.g., (a) certain comic book material and magazine material presentedly used for reading training may be justifiable from a motivational standpoint but may not contribute much knowledge relevant to future material that the student needs or wants to be able to read, (b) decoding rules and workbook materials usually have little impact upon knowledge and therefore are likely to be extremely inefficient with respect to increasing general reading ability.

3. be directed toward a specific subject matter area that either the student or the teacher deems as directly relevant to the type of reading material that the student is likely to encounter in the future.

4. not focus upon so called reading "skills" at the expense of gain in subject matter knowledge but instead should elevate gain in knowledge as a goal, at least until it equals if not exceeds the focus upon traditionally taught skills in reading.

REFERENCES

Anderson, Richard C. Control of student mediating processes during verbal learning and instruction. Review of Educational Research, 1970, 40, 349-370.

Bitzer, D. L.; Sherwood, B. A.; and Tenczar, P. Computer-based science education. University of Illinois, Computer-based Education Research Laboratory, May 1973.

Bormuth, J. R. Development of readability analyses. U.S. Office of Education Final Report, Proj. No. 7-0052, Contract Number OEC-3-7-070052-0326. University of Chicago, March 1969.

Carroll, J. B. A model of school learning. Teachers College Record, 1963, 64, 723-733.

Carver, R. P. A test of an hypothesized relationship between learning time and amount learned in school learning. The Journal of Educational Research, 1970, 64, 2, 57-58.

Carver, R. P. Reading as reasoning: Implications for Measurement. In W.H. Mac Ginitie (Ed.). Assessment problems in reading. Newark, Del.: International Reading Association, 1973, (pp. 44-56).

Carver, R. P. Revised procedures for developing reading-input materials and reading-storage tests. Journal of Reading Behavior, 1975, 2, 155-172.

Carver, R. P. Measuring prose difficulty using the Rauding Scale. Reading Research Quarterly, 1975-76, 11, 660-685.

Carver, R. P. Word length, prose difficulty, and reading rate. Journal of Reading Behavior, 1976, 8, 193-203.

Carver, R. P. Manual for the National Reading Standards. Kansas City, Mo.: Revrac Publications, 1977. (a)

Carver, R. P. Constancy of reading rate. Unpublished manuscript, 1977. (b)

Carver, R. P. Toward a comprehensive theory of reading and rauding. Reading Research Quarterly, 1978, in press.

Crano, W. D., Kenny, D. A. & Campbell, D. T. Does intelligence cause achievement?: A cross-lagged panel analysis. Journal of Educational Psychology, 1972, 63, 258-275.

Dore-Boyce, K., Misner, M.S., McGuire, L.D. Comparing reading expectancy formulas. The Reading Teacher, 1975, Oct., 8-14.

Eysenck, H. J. Check your own I.Q. Baltimore: Penguin, 1966.

Gagne, R.M. Expectations for school learning. AERA-Phi Delta Kappa Address, Annual meeting of the American Educational Research Association, New Orleans, 1973.

Geyer, J. J. Models of perceptual processes in reading. Reading Research Quarterly, 1973.

Guilford, J. P. The nature of human intelligence. New York: McGraw-Hill, 1967.

Harber, J. R. & Bryen, D.N. Black English and the task of reading. Review of Educational Research, 1976, 46, 387-405.

Jensen, A. R. How can we boost IQ and scholastic achievement? Harvard Educational Review, 1969, 39, 1-123.

Lohnes, P. R. & Gray, M.M. Intelligence and the Cooperative Reading Studies. Reading Research Quarterly, 1972, 7, 466-476.

Lykken, D. T. Statistical significance in psychological research. Psychological Bulletin, 1968, 70, 151-159.

Nelson, M.O. and Edelstein, G. Raven Progressive Matrices, Non-Language Multi-Mental Maturity Test, and California Test of Mental Maturity: Inter correlations. Psychological Reports, 1963, 13, 46.

Palermo, D. S. & Molfese, D. L. Language acquisition from age five onward. Psychological Bulletin, 1972, 78, 409-428.

Pezzullo, T. R.; Thorsen, E. D.; and Madans, G. F. The heritability of Jensen's Level I and II and divergent thinking. American Educational Research Journal, 1972, 9, 539-536.

Rohwer, W. D.; Ammon, M.S.; Suzuki, N.: and Levin, J.R. Population differences and learning proficiency. Journal of Educational Psychology, 1971, 62, 1-14.

Simon, H. A. & Chase, W.G. Skill in chess. American Scientist, 1973, 61, 394-403.

Singer, H. IQ is and is not related to reading. In S. Wanat (Ed.) Intelligence and reading. Newark, Del.: International Reading Association, 1976.

Smith, F. Understanding reading. New York: Holt, Rinehart, and Winston, 1971.

Smith, F. Psycholinguistics and reading. New York: Holt, Rinehart, and Winston, 1973.

Spache, G. D. Toward better reading. Champaign, Illinois: Garrard, 1963.

Spearman, C. Theory of a general factor. British Journal of Psychology, 1946, 36, 117-131.

Sticht, T. G. A program of Army functional job reading training: Development, implementation, and delivery system. Hum RRO-FR-WD. (CA)-75-7. Alexandria, Virginia: Human Resources Research Organization, June 1975.

Sticht, T. G. & Beck, L. J. Experimental Literacy Assessment Battery (LAB). Hum RRO-AFHRL-TR-76-51. Alexandria, Virginia: Human Resources Research Organization, August 1976.

Thorndike, E. L. Reading as reasoning: A study of mistakes in paragraph reading. Journal of Educational Psychology, 1917, 8, 323-332.

Thorndike, R. L. Reading as reasoning. Invited address at the annual meeting of the American Psychological Association, Washington, D.C., September 1971.

Vernon, P. E. The variations of intelligence with occupation, age, and locality. British Journal of Psychology, Statistical Section, 1947, 1, 5-63.

Vincent, D. V. The linear relationship between age and score of adults in intelligence tests. Occupational Psychology, 1952, 26, 243-249.

Appendix A

Rescaling the Raven to Measure Gain

Ronald P. Carver

University of Missouri-Kansas City

Abstract

The psychometrically developed Raven Progressive Matrices Test was rescaled so as to provide an interval scale that would be appropriate for use in situations where an edumetric test was needed. The scale was developed from a sample of 666 individuals ranging from Grade 3 to college students who had been administered the Raven. A Rescaled Raven scoring system was developed which provided, in effect, a weighted score based upon the relative difficulties of the items that were scored as correct. The Rescaled Raven scores were compared to the original scores in Grades 3-12, and the resulting functional relationships were interpreted as indicating that the Rescaled Raven does provide an interval scale and can be used to provide valid change and gain scores.

Rescaling the Raven to Measure Gain[1]

The primary purpose of this article is to present a method for transforming scores on the Raven Progressive Matrices Test so that they reflect an interval scale and can therefore be validly used to measure change, gain, or progress.

According to the test manual, the Raven test provides an index of an ability that is alternately described as "clear thinking," "intellectual capacity," and "reasoning." The Raven has been considered by Charles Spearman (1946), and others (see Vernon, 1947; Vincent, 1952), as the best non-verbal test for measuring _g_, general intelligence. A recently published study of the inheritance of general intelligence used the Raven as a measure of _g_, and found that about 85% of the variance in the _g_ trait, as estimated by this test, was inherited (Pezzullo, Thorsen, and Madaus, 1972). There is also interest in using the Raven in studies that do not focus upon individual differences, i.e., the Colored Raven was used by Rohwer, Ammon, Suzuki, and Levin (1971) to study group differences in the abstract thinking ability of children at different ages. Recently, Schonfield (1974) made inferences about the non-linear decline of intelligence with age using the Raven. If the Raven does not provide an interval scale, then the use of it in these latter situations could erroneously suggest non-linear functional relationships and erroneously suggest group-treatment interactions.

Since the Raven Progressive Matrices Test has not been designed to provide an interval scale, it seems unlikely that it would provide a scale that would be valid for reflecting gain, change, progress, or group differences.

Stated differently, the Raven has been designed as a psychometric test, not as an edumetric test. Psychometric tests have been designed to reflect differences between individuals under comparable treatment conditions whereas edumetric tests are designed to reflect group differences or changes within individuals as a result of different treatments (see Carver, 1974).

As stated earlier, the primary purpose of the present article is to present a technique for rescaling the Raven so that it does provide an interval scale. The result of this rescaling also will be evaluated using previously collected data.

Method and Results

Rationale. *One way to develop an interval scale would be to weight the items with respect to their relative difficulties. A more difficult item, i.e., one with a low _p_ value, would receive more weight than another less difficult item, i.e., one with a high _p_ value. If items were weighted in proportion to their _p_ values, then the resulting scale could be considered as interval; at least it should be more interval than the unweighted raw scores. The technique outlined above is the one that was used to rescale the Raven. The implimentation of this strategy for producing an interval scale was accomplished in four phases.

Phase 1. The Raven test had been administered to the 564 students in Grades 3-12 of a small town school system, in connection with the development of a reading test (see Carver, 1976a). Also, the Raven had been administered to 102 college students as part of a larger data collection effort (see Carver, 1976b). The _p_ values for each item, i.e., the percent who answered the item correctly, for these 666 individuals were calculated and are presented in Table 1.-

Insert Table I about here

Col. 1 of Table 1 contains the ranked difficulty of the 60 items. Col. 2 contains the identification label for each item, i.e., as given on the test booklets. Col. 3, as noted earlier, contains the p values for each item. Col. 4 contains the p values for the same 60 items as reported by Eisenberg, Levy, and Kirsch (1970). This tested population consisted of 402 adults, i.e., low-income laborers, college students and graduates, young adults confined for juvenile offenses, and workers being trained in a center for the mentally retarded. The correlation between the two sets of p values from the present sample and the adult sample was .98, signifying high consistency between the relative difficulties of the items in the two samples. From the data in Table 1, it can be seen that the Raven provides a wide range of p values with a large number clustered at the low end, i.e., there is a disproportionate high number of easy items. Thus, it would be difficult to justify considering the number correct score, i.e., the Raw Raven Score, as providing an interval scale.

Phase II. The Raven has a wide range of p values, producing a hierarchical scale. This means that if an individual answered a very difficult item, almost all items below that item in difficulty should also be answered correctly. To observe this effect, the items scores, 0 or 1, were reordered to reflect increasing item difficulty. In general, the lower difficulty items had scores of 1 while the higher difficulty items had scores of 0. Yet, there were deviations from this general effect. Some individuals with a raw score of 50, for example, would miss one or two of the extremely easy items. This was probably due to carelessness, either in marking an unintended response or by deciding too quickly c an answer. Other individuals who made extremely low scores,

15 for example, would often get some of the more difficult items correct. This was considered to be a chance effect since the test is a multiple-choice test and some items may be answered correctly by chance alone.

In order to eliminate the two above sources of unwanted error variance in the Raw Raven scores, a re-scoring system was developed. This system was developed so as to give credit for: (a) items that were answered incorrectly when they were easier items than other items that were answered correctly, and (b) items that were answered correctly when they were more difficult items than other items that were answered incorrectly. This system had as its goal the determination of the center of the transition set of items where the correctly answered items phased into the incorrected answered items. The following rules were used in the rescoring system:

Step 1. Record the score for each item, 0 for incorrect and 1 for correct, in ascending order of item difficulty using the item numbers and their corresponding p values given in Table 1. Table 2 contains an example--Col. 1 contains the Raven item label and Col. 2 contains the ranked item difficulty. Col. 3 contains the items scores, 0's and 1's, for example. In this example, the individual had a raw score of 24, i.e., 24 one's and 36 zero's in Col. 3.

Insert Table 2 about here

Step 2. Mark the item point corresponding to the total score, called Mark 1. In Col. 3, a line has been drawn after item 24 corresponding to the raw score of 24 and designating Mark 1.

Step 3. From Mark 1, count up 10 items and place another mark, Mark 2. From Mark 1 count 10 items below and place another mark, Mark 3. Notice in

Col. 3 that there are marks after Item 14 and Item 34, designating Mark 2 and Mark 3, respectively.

Step 4. If there are any 0's above Mark 2, they are changed to 1's, and if there are any 1's below Mark 3 they are changed to 0's. The results of Step 4 are presented for the example in Col. 4.

Step 5. Count the new number of 1's and place a mark after this item, Mark 4. In Col. 4, there are 20 1's so a mark has been placed after Item 20, i.e., Mark 4.

Step 6. From Mark 4, count up 3 items and place a mark, Mark 5. From Mark 4, count down 3 items and place another mark, Mark 6. Notice in Col. 4, there is a mark after item 17, Mark 5, and after Item 23, Mark 6.

Step 7. If there are any 0's above Mark 5, they are changed to 1's and if there are any 1's below Mark 6, they are changed to 0's. The results of this step are presented for the example in Col. 5.

Step 8. Count the new number of 1's and place a mark after this item, Mark 7. In Col. 4, there are 18 1's so a mark has been placed after Item 18, i.e., Mark 7.

Step 9. If there are any 0's above Mark 7, they are changed to 1's, and if there are any 1's below Mark 7, they are changed to 0's. The results of this step are presented for the example in Col. 6.

Step 10. Count the new number of 1's. This is the Rescored Raven score. The Rescored Raven score for the example in Table 2 is 18.

Counting up and down 10, 3, and 0 in Steps 3, 6, and 9, respectively, was decided arbitrarily.

For each of the Raven scores in Phase I, a Rescored Raven score was determined. So as to increase the accuracy and efficiency of this analysis, a computer program was written and used to derive the Rescored Raven scores.

New difficulty values, i.e., p values, were calculated for each item using the Rescored Raven scores. For each difficulty value, i.e., p, its complement was also calculated, i.e., $1-p$, and these "easiness" values were used for the rescaling procedures.

<u>Phase III</u>. Table 3 presents the Rescaled Raven scores for each Rescored Raven score. Col. 1 contains the identification label for each item, as described earlier. Col. 2 contains the Rescored Raven scores, as described in Phase II. Col. 3 contains the Rescored Raven easiness values, $1-p$, as determined from Phase II. These easiness values were considered as constituting the weight for each item, i.e., its relative worth. Since the Rescored Raven score is a perfect hierarchal scale, i.e., if Item 40 is correct then Items 1-39 must also be correct, then the total weighting associated with the Rescored Raven must consist of the cumulative weights of all items up to and including the most difficult item considered as correct. For example, the cumulative weight associated with a Rescored Raven score of 11, i.e., .10, was determined by summing the following item weights: .04, .02, .01, .01, .01, and .01. These cumulative weights for all 60 items are given in Col. 4 in Table 3. Finally, the relative weight for a given Rescored Raven score was determined by dividing each cumulative weight associated with that item by the total cumulative weight for all items and multiplying by 100. This has the effect of making the most difficult item worth 100 points and all other items worth less than 100 points. The last column in Table 3, Col. 5, contains the result of this procedure, i.e., these are the Rescaled Raven scores corresponding to each Rescored Raven score. For example, the Rescaled Raven score associated with a Rescored Raven score of 55 was 72 and it was determined by dividing the cumulative weight for the 55 score, 11.02, by the total cumulative weight for all 60 scores, 15.39. The Rescaled Raven score should provide an interval scale since it represents, in effect, the percent of the total difficulty of the test that is represented by each successively higher score on the test.

Insert Table 3 about here

Phase IV. Table 3 may be used to convert a Rescored Raven score into a Rescaled Raven score. However, there are occasions where only a Raven Raw score is available, e.g., from data where the original answer sheets are not available so that the Rescored Raven score could not be determined. In these situations, it would be desirable to be able to get an estimated Rescaled Raven score given only the Raven Raw Score. The method used to derive estimated Rescaled Raven scores will now be described in detail.

Step 1. For the original 666 tests, the frequency distribution of the Rescored Raven scores for each Raven Raw score from 0 to 60 was determined.

Step 2. The median Rescored Raven score was determined for each Raven Raw score.

Step 3. These median values were transformed into Rescaled Raven scores using Table 4. There was a certain amount of undesirable variance in this result, i.e., the curve that related the Raw scores to the Rescaled scores was not perfectly monotonic.

Step 4. So as to provide a transformation that was likely to be more reliable, the curve determined from Step 3 was smoothed twice, i.e., the mean was calculated for each adjacent pair of Rescaled scores once, and then this procedure was repeated. The results from Step 4 are contained in Table 4, and are presented graphically in Fig. 1.

Insert Table 4 and Fig. 1 about here

Notice in Table 4 that there were no subjects who scored between 1 and 6. Parentheses were placed around the lowest 0 and the highest 100 because these values were assumed instead of empirically determined; when a curve is smoothed twice the two end data points are lost in the process. Notice in Fig. 1 that there is a positively accelerated relationship between the estimated Rescaled Raven scores and the actual Raw Raven scores. For example, at the lower end of the Raw Raven scale a 10 point increase from 15 to 25 would reflect only a 4 point increase in Rescaled Raven scores, 1 to 5, whereas a similar 10 point increase in Raw Raven scores from 45 to 55 would result in a 34 point increase in Rescaled Raven scores, 51 to 85.

Evaluating the Rescaled Raven Scores

Fig. 2 contains the 25th, 50th, and 75th percentile scores for the Raw Raven for each of Grades 2-12 in the original sample. Notice that there is an initial high acceleration which decreases rapidly as the grade in school increases. These data could be interpreted, erroneously, as indicating that general intelligence increases rapidly in the early school grades and then does not change much in the latter years of school. Furthermore, the difference between the 25th and 75th percentile scores at Grade 2 is about twice as large as it is at Grade 12 and this could be interpreted, erroneously, as indicating that the amount of variance in general intelligence decreased as children got older, i.e., from Grade 2-Grade 12.

Insert Fig. 2 about here

Fig. 3 contains the relationship between the Rescored Raven scores and school grade, i.e., the same type of relationship as is presented in Fig. 2

except Rescored values were used instead of Raw values. Notice that the nature of the relationship in Fig. 3 is almost exactly the same as in Fig. 2 except the higher Rescored scores tend to be higher than the corresponding Raw scores and the lower Rescored scores tend to be lower than the corresponding Raw scores.

Insert Fig. 3 about here

Fig. 4 contains the *25th, 50th, and 75th percentile Rescaled Raven* scores for each of Grades 2-12. This is the same type of relationship as is presented in Fig. 2 and Fig. 3 except in this case there is a linear increase in these scores with each increase in grade for all three of these percentile scores. Notice also that the variability in the scores tends to be relatively constant, i.e., the interquartile range does not tend to be twice as large at Grade 2 as it is at Grade 12. The regularity and linearity of these three curves suggests that the Rescaled Raven scores probably do produce a reliable and valid interval scale.

Insert Fig. 4 about here

Implications

It appears that the rescaled Raven does produce a scale that has interval properties. At least there are good rational and empirical grounds for purporting that it provides much more of an interval scale than the raw scores. Thus, any time that the Raven is used to measure gain, change, progress, group differences, or treatment differences, it would seem that it would be best to use the Rescaled scores because an interval scale would be best in these situations. One of the dangers involved in not using the rescaled scores is that an *artifactual interaction could result. For example,* suppose the scores of Low SES second graders and third graders are compared to the scores of High SES second

graders and third graders. Fig. 5 contains some hypothetical data to illustrate this problem. Notice that when the Raw Raven scores are plotted as a function of grade in school there is an interaction effect suggesting, erroneously, that the difference between socioeconomic groups decreases from second grade to third grade. Notice that when these Raw Raven scores are transformed into *estimated Rescaled Raven* scores, using Table 4, there is no interaction indicated.

Insert Fig. 5 about here

The above problem, *illustrated* in *Fig.* 5, has *implications for the current* popular focus upon aptitude-treatment interactions in educational research (e.g., see Cronbach, 1975). It would seem that whenever an aptitude-treatment interaction is found that it would behoove the investigator to rule out the alternative hypothesis that the interaction was simply an artifact of the measurement scale. The likelihood of this spurious interaction effect is increased because differences in aptitude suggest that there will be differences compared from the extremes of the scale. If the scale is not an interval scale, it would seem to be just as likely that such interactions are artifactual as it is likely that they are real. Since most psychometric measurement devices have not been designed to provide an interval scale and do not purport to provide an interval scale, it would seem better to refrain from studying or interpreting interactions using these devices.

A *final* implication of the Rescaled Raven data for grades 2-12 is that growth in general intelligence, at least as reflected by the Raven, is relatively constant for all levels of intellectual ability. The bottom half of each class (as *indicated by the 25th percentile scores*) showed just as much growth each year as the top half of the class (as indicated by the 75th percentile score). These data also could be interpreted as suggesting that the bottom half of the class would be ready to undertake the same level of intellectual challenges about 4-5 years later than the top half of each class. Stated differently, if the top half of each class, intellectually, would have entered Grade 1 at age 4 and the bottom half of each class had entered Grade 1 at age 8, then the grades might have been better grouped for intellectual challenges throughout all 12 grades since their rates of intellectual growth appear to be approximately equal.

In conclusion, it would appear to be difficult to justify using the *raw* scores on the Raven when an interval scale is desired. The rescaled Raven scores should be used, at least until a better scale is developed. Furthermore, this rescaling of the Raven provides a concrete illustration of how a test that is designed to reflect individual differences, i.e., a psychometrically designed test, may produce misleading or invalid results when it is used to reflect change, gain, or progress, i.e., situations where an edumetrically designed test should be used.

References

Carver, R. P. Two dimensions of tests: Psychometric and edumetric. American Psychologist, 1974, 29, 512-518.

Carver, R. P. *Manual for the National Reading Standards*. Kansas City, Mo.: Revrac Publications, in press, 1976. (a)

Carver, R. P. Optimal rate of reading. Unpublished manuscript, 1976. (b)

Cronbach, L. J. Beyond the two disciplines of scientific psychology, American Psychologist, 1975, 30, 116-127.

Eisenberg, B. S., Levy, B. I., & Kirsch, A. D. Some psychometric properties of the Raven Progressive Matrices Test. Washington, D. C.: American Psychological Association, Experimental Publication System, Ms. No. 230-2, Issue No. 7, August 1970.

Pezzullo, T. R., Thorsen, E. E., & Madaus, G. F. The heritability of Jensen's Level I and II and divergent thinking. American Educational Research Journal, 1972, 9, 539-546.

Rohwer, W. D., Jr., Ammon, M. S., Suzuki, N., & Levin, J. R. Population differences and learning proficiency. Journal of Educational Psychology, 1971, 62, 1-14.

Schonfield, D. Translations in gerontology--from lab to life. American Psychologist, 1974, 29, 796-801.

Spearman, C. *Theory of a general factor.* British Journal of Psychology, 1946, 36, 117-131.

Vernon, P. E. The variations of intelligence with occupation, age, and locality. British Journal of Psychology, Statistical Section, 1947, 1, 52-63.

Vincent, D. V. The linear relationship between age and score of adults in intelligence test. Occupational Psychology, 1952, 26, 243-249.

Footnote

1 The preparation of this article was supported in part by the U. S. Office of Naval Research, Personnel and Training Research Programs, Contract No. 0014-72-C-0240, and in part by the National Institute of Education, Grant No. NE-G-00-3-0043.

Table 1

Difficulty Indices for the 60 Raven Test Items From Two Different Samples

Col. 1 Difficulty Rank	Col. 2 Raven Item	Col. 3 Present Sample (N=666) p	Col. 4 Adult Sample (N=402) p	Col. 1 Difficulty Rank	Col. 2 Raven Item	Col. 3 Present Sample (N=666) p	Col. 4 Adult Sample (N=402) p
1	A1	1.000	.99	31	C 4	.737	.82
2	A2	.997	.99	32	A11	.734	.85
3	A3	.994	.99	33	D 7	.710	.80
4	A4	.994	.99	34	B 8	.694	.73
5	A5	.988	.98	35	C 9	.691	.72
6	A6	.986	.97	36	D10	.689	.77
7	B1	.986	.97	37	D 8	.685	.77
8	B2	.982	.95	38	E 1	.674	.77
9	A9	.946	.94	39	C 6	.662	.74
10	D1	.943	.94	40	B 7	.658	.74
11	B3	.934	.96	41	E 2	.631	.76
12	C1	.931	.93	42	D 9	.623	.75
13	A8	.916	.89	43	C 8	.599	.75
14	A10	.905	.88	44	E 3	.566	.72
15	C2	.904	.90	45	A12	.560	.66
16	A7	.902	.89	46	E 5	.548	.70
17	D5	.883	.88	47	B12	.538	.62
18	D2	.875	.86	48	E 4	.503	.68
19	B4	.871	.89	49	C10	.476	.66
20	D3	.850	.85	50	E 7	.465	.58
21	C3	.848	.88	51	E 6	.453	.61
22	C5	.842	.85	52	C11	.416	.58
23	B10	.826	.83	53	E 8	.390	.56
24	B5	.818	.85	54	D11	.300	.39
25	D4	.817	.85	55	E 9	.281	.52
26	D6	.809	.86	56	E10	.207	.36
27	C7	.767	.82	57	C12	.191	.30
28	B9	.766	.80	58	D12	.189	.30
29	B11	.758	.75	59	E12	.156	.23
30	B6	.755	.82	60	E11	.084	.14

Table 2

An Example of how the Rescored Score Is Obtained

Col.1 Raven Item	Col.2 Diff. Rank	Col.3 Item Score	Col.4	Col.5	Col.6 Rescored Score	Col.1 Raven Item	Col.2 Diff. Rank	Col.3	Col.4	Col.5	Col.6 Rescored Score
A 1	1	1	1	1	1	C 4	31	0	0	0	0
A 2	2	1	1	1	1	A11	32	0	0	0	0
A 3	3	1	1	1	1	D 7	33	0	0	0	0
A 4	4	1	1	1	1	B 8	34	0	0	0	0
A 5	5	1	1	1	1	C 9	35	0	0	0	0
A 6	6	1	1	1	1	D10	36	0	0	0	0
B 1	7	1	1	1	1	D 8	37	0	0	0	0
B 2	8	1	1	1	1	E 1	38	1	0	0	0
A 9	9	1	1	1	1	C 6	39	0	0	0	0
D 1	10	1	1	1	1	B 7	40	0	0	0	0
B 3	11	1	1	1	1	E 2	41	0	0	0	0
C 1	12	1	1	1	1	D 9	42	0	0	0	0
A 8	13	1	1	1	1	C 8	43	1	0	0	0
A10	14	1	1	1	1	E 3	44	0	0	0	0
C 2	15	1	1	1	1	A12	45	0	0	0	0
A 7	16	1	1	1	1	E 5	46	0	0	0	0
D 5	17	1	1	1	1	B12	47	0	0	0	0
D 2	18	0	0	0	1	E 4	48	0	0	0	0
B 4	19	0	0	0	0	C10	49	0	0	0	0
D 3	20	1	1	1	0	E 7	50	0	0	0	0
C 3	21	0	0	0	0	E 6	51	0	0	0	0
C 5	22	0	0	0	0	C11	52	0	0	0	0
B10	23	0	0	0	0	E 8	53	1	0	0	0
B 5	24	0	0	0	0	D11	54	0	0	0	0
D 4	25	1	1	0	0	E 9	55	0	0	0	0
D 6	26	0	0	0	0	E10	56	0	0	0	0
C 7	27	1	1	0	0	C12	57	0	0	0	0
B 9	28	0	0	0	0	D12	58	1	0	0	0
B11	29	0	0	0	0	E12	59	0	0	0	0
B 6	30	0	0	0	0	E11	60	0	0	0	0
							Total	24	20	18	18

Table 3

Determining the Rescaled Raven Scores from the Rescored Raven Scores

Col.1 Raven Item	Col.2 Rescored Raven	Col.3 Phase II Easiness Values (Weights) 1-p	Col.4 Cumul-ative Weights	Col.5 Rescaled Raven	Col.1 Raven Item	Col.2 Rescored Raven	Col.3 Phase II Easiness Values (Weights)	Col.4 Cumul-ative Weights	Col.5 Rescaled Raven
A 1	1	.00	.00	0	C 4	31	.17	2.16	14
A 2	2	.00	.00	0	A11	32	.18	2.34	15
A 3	3	.00	.00	0	D 7	33	.18	2.52	16
A 4	4	.00	.00	0	B 8	34	.19	2.71	18
A 5	5	.00	.00	0	C 9	35	.20	2.91	19
A 6	6	.01	.01	0	D10	36	.20	3.11	20
B 1	7	.01	.02	0	D 8	37	.21	3.32	22
B 2	8	.01	.03	0	E 1	38	.22	3.54	23
A 9	9	.01	.04	0	C 6	39	.24	3.78	25
D 1	10	.02	.06	0	B 7	40	.25	4.03	26
B 3	11	.04	.10	1	E 2	41	.27	4.30	28
C 1	12	.04	.14	1	D 9	42	.30	4.60	30
A 8	13	.04	.18	1	C 8	43	.31	4.91	32
A10	14	.05	.23	2	E 3	44	.34	5.25	34
C 2	15	.06	.29	2	A12	45	.36	5.61	36
A 7	16	.06	.35	2	E 5	46	.40	6.01	39
D 5	17	.07	.42	3	B12	47	.43	6.44	42
D 2	18	.08	.50	3	E 4	48	.47	6.91	45
B 4	19	.09	.59	4	C10	49	.50	7.41	48
D 3	20	.09	.68	4	E 7	50	.53	7.94	52
C 3	21	.11	.79	5	E 6	51	.56	8.50	55
C 5	22	.11	.90	6	C11	52	.58	9.08	59
B10	23	.12	1.02	7	E 8	53	.61	9.69	63
B 5	24	.12	1.14	7	D11	54	.65	10.34	67
D 4	25	.13	1.27	8	E 9	55	.68	11.02	72
D 6	26	.13	1.40	9	E10	56	.74	11.76	76
C 7	27	.14	1.54	10	C12	57	.82	12.58	82
B 9	28	.14	1.68	11	D12	58	.90	13.48	88
B11	29	.15	1.83	12	E12	59	.94	14.42	94
B 6	30	.16	1.99	13	E11	60	.97	15.39	100

Table 4

Rescaled Raven Scores Estimated from Raw Raven Scores

Col. 1 Raw Raven	Col. 2 Estimated Rescaled Raven	Col. 1 Raw Raven	Col. 2 Estimated Rescaled Raven	Col. 1 Raw Raven	Col. 2 Estimated Rescaled Raven
1	-	21	3	41	37
2	-	22	4	42	38
3	-	23	4	43	41
4	-	24	5	44	46
5	-	25	5	45	51
6	-	26	6	46	55
7	(0)	27	8	47	59
8	0	28	10	48	64
9	0	29	12	49	68
10	0	30	15	50	72
11	0	31	17	51	75
12	0	32	18	52	78
13	0	33	20	53	80
14	1	34	24	54	83
15	1	35	26	55	85
16	2	36	28	56	85
17	2	37	30	57	88
18	2	38	32	58	94
19	2	39	34	59	100
20	3	40	36	60	(100)

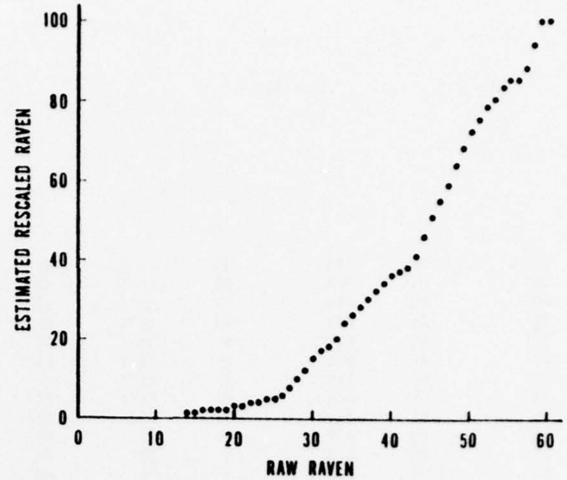


Fig. 1. Estimated Rescaled Raven scores as a function of Raw Raven scores.

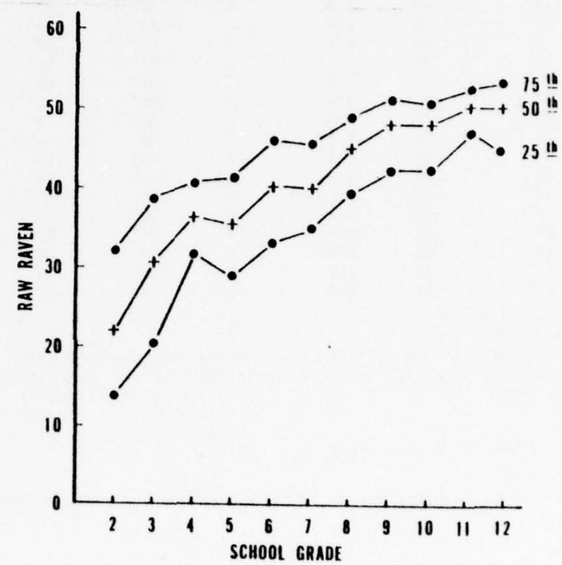


Fig. 2. The 25th, 50th and 75th Raw Raven percentiles for each school grade.

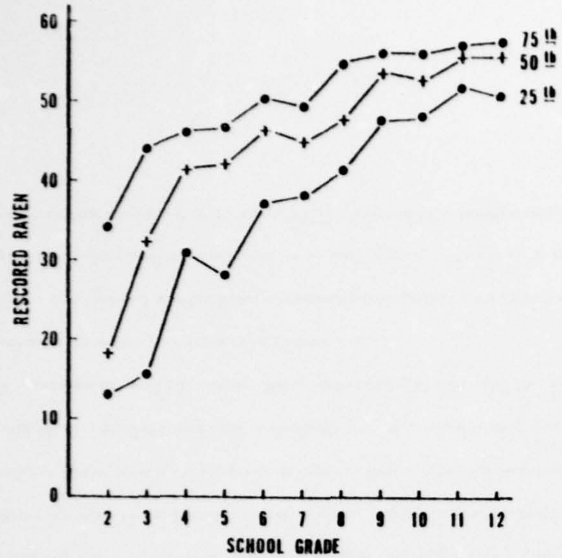


Fig. 3. The 25th, 50th, and 75th Rescored Raven percentiles for each school grade.

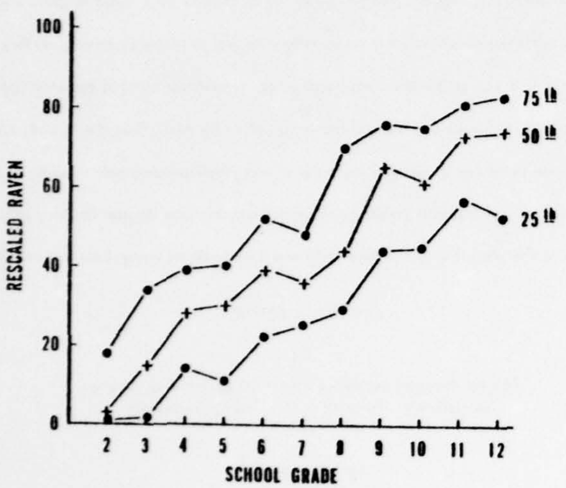


Fig. 4. The 25th, 50th, and 75th Rescaled Raven percentiles for each school grade.

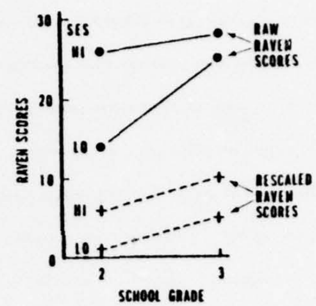


Fig. 5. An illustration of how an interaction effect using Raw Raven Scores is artifactual, i.e., there is no interaction when these scores are transformed into Rescaled Raven scores.

Appendix B

Intellectual Ability and Reading: Comparing Actual Reading Ability with Expected Reading Ability

by

Ronald P. Carver

Reading expectancy formulas generally use some measure of intelligence to estimate the grade level of reading ability that could be expected from an individual. For example, the Harris (1970) formula uses chronological age and mental age while the Bond and Tinker (1957) formula uses "years in school" and IQ to get an expected reading grade level. These formulas do not specify how IQ is to be measured. Most often the measure of intelligence requires either reading or the use of words so the intelligence measure is measuring some of the same ability as the reading test. Conversely, it has been pointed out that the questions on reading tests often require so much inference or reasoning on the part of the reader that it is questionable whether the reading test is primarily measuring reading or intelligence (see Carver, 1975).

The primary purposes of the present research wereto: (a) develop a reading expectancy formula which was based upon a measure of intelligence that was not contaminated by word knowledge and which was also based upon a reading test that was not unduly contaminated by intelligence requirements, and (b) compare the actual reading ability with the expected reading ability of groups of good and poor readers in Grades 2-12.

Recently, a standardized test has been developed which provides a measure of reading ability which seems to be relatively independent of intelligence, inference, or reasoning. It is a modified cloze type of procedure (Carver, 1977). Instead of deleting every fifth word,

as is the case with the regular cloze task, a word is added every fifth word so the task is to choose which of the two words belongs (Carver, 1975). This task reduces inference and other "intelligence" and reasoning type tasks to a minimum. Furthermore, this test claims to provide an interval scale for measuring reading progress. Most standardized tests are developed using the traditional psychometric methods (see Carver, 1972; Carver, 1974), and do not even purport to provide an interval scale for measuring change, gain, or progress. However, these tests are often used as if they did provide a valid interval scale. With an interval scale, one can validly compare differences, e.g., differences between good and poor readers at Grade 2, as compared to Grade 12.

As for measuring intelligence independent of reading or verbal ability, the Raven Progressive Matrices Test seemed to provide such a measure. The Raven has been considered by Spearman (1946), and others (Jensen, 1969; Vincent, 1952) as providing a relatively pure measure of intelligence. It involves 60 figures, or what might be called puzzles with missing parts,and the task is to decide which of several alternatives is the missing part for each of the 60 items. Recently, the scores on this test were rescaled to provide an interval scale for measuring change, gain, or progress in reasoning or intellectual ability (See Appendix A).

In Phase I of the following research, an expectancy formula was developed which used scores on the Raven test to predict reading achievement as measured by the aforementioned reading test.

In Phase II of this research, actual reading ability was compared to expected reading ability. Earlier research using the Raven in Grades 2-12 of a small town, rural school system indicated that the scores from the top half of each gradewere about 4-5 years higher than the bottom half of each grade (See Appendix A). This was interpreted as suggesting that the top half of the class could start to school at age 4 and the bottom half at age 8 and this would tend to make the classes more homogeneous, intellectually. Since the top half of the class does not start until age 6, it was hypothesized that the top half of each grade would not be achieving up to their potential, i.e., their actual scores would be lower than their potential scores. However, as they progressed through the grades their actual scores should tend to approach their expected scores. As for the lower half of the class, it was hypothesized that they would drop further and further below their potential as they progressed through school. The instruction would tend to be difficult for them and they would be unable to achieve mastery at each level and would fall further and further behind due to a compounding of unmastered skills and a negative attitude toward reading (see Bond & Tinker, 1957).

Phase I

Introduction

The students in Grades 1-12 of a small school system had been administered the National Reading Standards test on a previous occasion as part of another project (see Carver, 1977). Within the same week, the Raven was administered. The general strategy for determining the expected reading score involved the determination of the average reading test score and the average Raven test score in each grade.

Method

Subjects. The scores from approximately 400 students in Grades 2-12 were used in this research. All students who were attending school on any particular testing day were tested. The school system is a small town, rural school located in the midwest. On other standardized

tests, the average score for students in each grade is generally near the 50th percentile. The school system is mostly traditional with around 25-35 students in a regular classroom, with a single classroom teacher in the elementary grades. In Junior High and Senior High, the students attend subject matter classes in different classrooms with different teachers throughout the day.

Reading Test. The reading test was the National Reading Standards. It has five levels and two forms, A and B, at each level. Each higher level includes more passages and increasingly difficult passages, while the time limit remains at 15 minutes for all levels and forms of the test. The primary score on the test is the Reading Efficiency (RE) score. It is determined from the raw scores and it has been developed so as to provide an interval scale from 0 up to 169. The Reading Efficiency, RE, score may be transformed into grade level scores varying from Grade 1 to Grade 16. The grade level scores are not perfectly linearly related to the RE scores through the entire range.

The National Reading Standards (NRS) is a sequential type test, i.e., different levels of the test are administered until the test is found which is most appropriate for the level of ability of the student. For example, if an individual scores toward the top of Test 3, then Test 4 is administered. For the NRS test, two scores were derived by administering two or more From A tests and two or more Form B tests. For this research, the mean of the two Reading Efficiency scores for each person, from Form A and Form B, was used.

Intelligence Test. As noted earlier, the scores on the Raven Progressive Matrices Test were used to estimate potential reading ability. There are 60 items on the test but the items vary greatly in difficulty, and there are more easy items than difficult items. A new scoring and scaling procedure has been developed which provides Rescaled Raven Scores that range from 0-100 on an equal interval basis (See Appendix A). The Rescaled Raven score was determined for each individual who took this test and this was the score that was analyzed.

Statistical Procedures and Results. In each of the 11 grades, i.e., Grades 2-12, the median Rescaled Raven score for the grade was determined and the corresponding median Reading Efficiency, RE, score for the grade was also determined. For these eleven data points, a least squares straight line was fit and the following prediction equation resulted:

$$\text{Estimated RE Score} = 1.14 \times \text{Rescaled Raven Score} + 9.94 \qquad (1)$$

The above equation was then used to get the expected RE scores for all the Rescaled Raven scores, as listed in Table 1. In Table 1 the Rescaled Raven scores are symbolized as $\hat{I}_a$, i.e., an estimate of intellectual ability, and the expected Reading Efficiency score is symbolized as $\widehat{RE}$. Therefore, Equation I may be rewritten as:

$$\widehat{RE} = 1.14\ \hat{I}_a + 9.94 \qquad (2)$$

Also in Table 1 is a third column headed by the symbol $\hat{G}_a$. The symbol $\hat{G}_a$ has been used to designate the grade level of ability that is associated with each $\widehat{RE}$ score. The expected grade levels of ability, $\hat{G}_a$, in Table 1 have been determined by using the tables developed by Carver (See Appendix A) for converting RE scores into G_a scores. Notice that the grade level equivalents, G_a, for the Rescaled Raven scores, $\hat{I}_a$, range from about Grade 2 to about Grade 14.

Insert Table 1 about here

Discussion. Table 1 was used to convert Rescaled Raven scores into Expected Reading Ability scores in the Phase II research which will be presented later. It should be noted that

Table 1 could be used by any researcher or practitioner to investigate and compare expected reading ability, $\widehat{RE}$, with actual reading ability, RE. First, the Raven would need to be administered and a Rescaled score determined using the procedures outlined elsewhere (See Appendix A). Given the Rescaled score, the expected reading ability, $\widehat{RE}$, could be determined from Table 1. Then, the National Reading Standards test could be administered to determine the actual reading ability, RE.

It should be pointed out that the expected reading ability scores are based upon the average intellectual ability and average reading achievement of the students in this particular school system. That is, if Table 1 is used by other researchers or practitioners to calculate reading expectancy, then the expectancy is based upon this school system as an anchor school system for comparison purposes. Since the students in this school system tend to score average or near average on other nationally standardized tests, it seems reasonable to use this school system as an anchor or bench mark for comparing other individuals or school systems. Yet, it should not go unnoticed that any comparisons between actual and expected scores are based upon the gains of the average students in this particular school system.

The type of normative comparison discussed above is not traditional in testing. Psychometricians have advocated the collection of national samples for comparison purposes. However, any group may be used as a standard or reference group for interpretation purposes as long as it is not forgotten that the scores derive their meaning in terms of this anchor group.

Phase II

Introduction

In Phase II, the actual reading ability, RE, was compared to the expected reading ability, $\widehat{RE}$, in the top and bottom halves of each of Grades 2-12. As mentioned earlier, it was hypothesized that the top half of the class would tend to get closer and closer to their potential or expected reading ability as they progressed through the grades while the bottom half of the class would tend to drop further and further below their potential or expected reading ability.

Statistical Procedures. In each grade, the top half of each class was represented by its middle score, i.e., the 75th percentile, and the bottom half was similarly represented by its middle score, i.e., the 25th percentile. Therefore, for the RE scores in each grade, the 75th and 25th percentiles were determined. For the $\hat{I}_a$ scores in each grade, the 75th and 25th percentiles were also determined. These latter $\hat{I}_a$ scores were converted into $\widehat{RE}$ scores using Table 1.

An alternate statistical analysis which might seem to be more desirable would involve determining the top and bottom half of the class based upon the RE scores and then getting the median Raven scores of these two groups. Or, vice versa, determine the top and bottom half of the class on the basis of the Raven scores and then get the median RE scores for these two groups. As desirable as these two solutions might seem to be, they both suffer from regression to the mean effects and are, therefore, likely to give biased estimates.

Results and Discussion. Fig. 1 contains the actual reading scores, RE, and expected reading scores, $\widehat{RE}$, for the top half of the class, 75th percentile, and the bottom half of the class, 25th percentile.

It may be noted that the top half of the class started out reading at a level that was lower than their expected level, i.e., in Grades 2-4. From Grades 5-9, they read at levels equal to what might be expected from the scores on the Raven. However, in grades 10-12, they read at levels that were higher that what might be expected. This result was somewhat as expected in that the top half of the class did well in relationship to their expectancy level. However, the scores from the bottom half of the class were not as expected. There appears to be no tendency for those in the lower half of the class to do worse than what might be expected from their estimates of intellectual capacity or reasoning ability. This latter result is somewhat surprising in that it was expected that the lower half of the class would not be working up to their potential and, therefore, the actual scores would tend to fall further below the expected scores as the students progressed through school.

Insert Fig. 1 about here

Another aspect of the data in Fig. 1 concerns the spread or the amount of difference between the bottom half of the class as compared to the top half of the class with respect to actual reading ability, RE. Notice that the bottom of the class is not as far behind the top half of the class in the lower grades as it is in the higher grades. The spread between the top and bottom halves tends to increase as the grades in school increase. In Grade 2 the difference in the RE scores was about 10 and in Grade 12 the difference was about 36. This is not an unusual finding, i.e., it is fairly common to find that the lower ability students tend to do worse in relationship to the higher ability students as the grade in school increases. However, this finding has often been interpreted as indicating that something was defective about the quality of the schooling of the lower ability individuals so that they were not achieving up to their potential. These data do not support this idea. In fact, these data suggest that *the poorer readers in each grade are reading at a level that is commensurate with their general intellectual ability*. The increase in the spread between the bottom of the class and the top of the class seems to be due solely to the fact that the top half of the class does not read

as well as could be expected in the lower grades, catches up to their expected ability in the middle grades and then exceeds their expected ability in the upper grades. Thus, the increase in the differences between the top and bottom half of the class seems to be due to the deviate performance of the top half of the class. The bottom half of the class does not seem to deviate in that they are achieving at a rate that would be expected from their general intellectual ability.

After the above data had been analyzed, a question arose as to whether these general results would be the same for *individuals who were even higher in reading ability* and also for individuals who were even lower in reading ability. Therefore, similar actual and expected reading ability scores were determined for the 95th percentile and the 5th percentile. These percentiles could be regarded as *representing the top 10% of the class and the bottom 10% of the class*. These data are presented in Fig. 2.

Insert Fig. 2 about here

Also included in Fog. 2 are the actual and expected scores for the class as a whole, i.e., the 50th percentile. The 50th percentile scores, i.e., the median, were the scores from which the expected scoring system was derived in Phase I, i.e., the expected scores were developed so that they would match as closely as possible the actual scores at the 50th percentile. Therefore, the high degree of correspondence between the actual and expected scores for the class as a whole, i.e., *the 50th percentile, is not a research finding but a consequence* of the way the expected scores were derived.

The data for the top 10% and bottom 10% of the class tend to replicate the earlier results except the results for the top 10% are even more pronounced. That is, the top 10% tends to read further below what would be expected at the lower grades and then by the higher grades they have accelerated in their ability so rapidly that they read considerably above what would be expected.

For the bottom 10% of the class, the correspondence between the actual performance and the expected performance is extremely close, possibly even closer than it was with the bottom half of the class in Fig. 1. In fact, the degree of correspondence between actual and expected scores for the *bottom* 10% of the class is about equal to that of the class as a whole, i.e., the 50th percentile data which was derived to provide high correspondence. Even the seemingly large difference that exists between the actual and expected scores in Grades 2 and 3 is artifactual. The expected scores cannot go below 10, i.e., the 10 value is not accurate but actually indicates that the expected score is 10 or below.

These data in Fig. 2 further support the earlier results. There is no evidence that the poorer readers are not reading up to their potential. There is evidence that the better readers are not reading up to their potential in the elementary school years.

It might be helpful to discuss the results in Fig. 2 using the Grade Equivalent values, i.e., $\hat{G}_a$, given in Table 1. The top 10% of the class in Grade 2 reads extremely well, *i.e., at about the sixth grade level,* $\hat{G}_a = 5.5$. *However, they might be expected to be* reading at about the ninth grade level, $\hat{G}_a = 8.7$; they are reading about 3 grade levels below what might be considered as their potential. The bottom 10% of the class does not reach the sixth grade level of reading ability until they are in Grade 10, i.e., $G_a = 6.0$,

and they are reading at a level almost precisely equal to what might be considered as their potential, i.e., $\hat{G}_a = 6.2$. In this example, it can be seen that the bottom 10% of the class was about eight years behind the top 10% of the class, i.e., the 10th graders in the bottom 10% achieved as much reading ability as the 2nd graders in the top 10% achieved. Furthermore, the top 10% of the class of second graders do not seem to be reading as well as they are capable, whereas the bottom 10% do seem to be reading as well as they are capable.

Intellectual Ability and Reading Ability

Before the implications of these results are discussed in the next section, certain crucial assumptions need to be outlined, i.e., the assumptions dealing with intelligence and reading. *It is assumed* that the Raven test provides a good estimate of potential or expected reading ability. The Raven has been categorized by Guilford (1967) in his model of *intelligence* as a measure of the Cognition of Relations between Figures. Reading involves the Cognition of the Relations between Words. Thus, only the mode of response is different, Figures versus Words, since both require Cognition of Relations.

It is further assumed that reading is a type of intellectual activity which can be learned or taught under certain favorable learning circumstances, e.g., reading instruction. However, it is assumed that it is not easy to teach the type of reasoning skills that are required on the Raven, so that it reflects a basic intellectual ability that is not readily amenable to change. In this regard, Pezzulo, Thorsen, and Madans (1972) interpreted the scores on the Raven as being 85% inherited. It is assumed that scores on the Raven would increase each year during the school years, due to a maturation effect, even if the students did not attend school. However, it is assumed that scores on a reading test such as the NRS used in this research, would not increase each year during the school years if the students did not receive the kind of learning experiences expected from attending school. Thus, the Raven is used as an index of the level of the student's reasoning ability and it is assumed that it is reasonable to expect a student's level of reading ability to be equal to his/her level of reasoning ability. Stated differently, it seems unreasonable to expect a student (e.g., in Grade 10) who reads poorly in relationship to his same age peers (e.g., Grade 2 reading level) to be able to read as well as his same age peers when his ability to reason is also very poor in relationship to his same age peers (e.g., Grade 2 intellectual level).

It is being assumed that a student whose intellectual ability is equal to an average second grader should not be expected to be able to read any better or any worse than an average second grader, no matter what grade the student is in. In fact, if a tenth grader does fit the above description, then it seems unfair to the student, unfair to the teachers, unfair to the parents, and unfair to the school system to categorize him/her as a "poor" reader. This student would not be considered a poor reader in relationship to his expected achievement level in reading.

Most of the above assumptions about the relationship between intellectual ability and reading ability are no different from those ordinarily made by those involved in reading. However, *it seemed necessary to make them explicit as a review prior to discussing the implications* of the research.

Implications of the Research

For the better readers in the early grades, these results suggest that something might be done to help them read as well as they can reason. They do not seem to be reading as well as they could be reading. Probably, these children could achieve at their expected level if

they were given the proper learning experiences in preschool, nursery school, kindergarten, and the early school years. The present data might be interpreted as suggesting that many school systems are not providing an instructional environment that allows the high potential *of its better readers to be realized. On the other hand, the later high success of the better* readers, i.e., they read better than might be expected, could be attributable to the present instructional system. Maybe the reason why the good readers later become over achievers is that they had so much early success that they were highly motivated to read and learn more. Therefore, because of the extra practice and background knowledge they acquired, they eventually exceeded all prior predictions regarding reading ability. If this is a reasonably valid explanation for the high success of the good readers in the upper grades, then any effort to introduce reading at the preschool levels should be careful not to discourage reading by going too far, too fast. Introducing too many failures or negative learning experiences early may be counterproductive in the long run.

The implications for the poorer readers in each grade are more crucial. These data suggest that, as a group, the poorer readers in each grade are reading at their potential. At least, they are reading as well for their general reasoning ability as the average readers in each grade. Thus, special programs to help these students as a group read better may be ill advised. The introduction of special reading programs for these students suggests that something is "wrong" with them when these data would suggest that there was nothing "wrong" with them or their instruction at all. Just because they read very poorly in relationship to their same age peers does not necessarily mean that they should be singled out for special instruction. Only if they read poorly in relationship to their estimated potential reading ability should they receive special attention. Furthermore, it seems fruitless to expect these students to make large gains in reading ability without making large gains in their general ability to reason. This same conclusion has also been drawn by Thorndike (1971). However, it seems unlikely that school systems can have much of an effect upon a student's general ability to reason.

Teachers and school systems are often criticized and held accountable for the low reading ability of many students. In this research, for example, around 10% of the 10th graders were reading at the second grade level. This could be interpreted as a condemnation of the school system with respect to these students. Yet, it may be remembered that the lower 10% of the 10th grade class was reasoning at a level that was also equal to the second graders. With this information, it seems unfair to hold the school system accountable for this performance of its students. In fact, the school system might be commended since the instruction for these students appears to have been commensurate with that of the instruction for the average students in the class. If any group got shortchanged by the system, it might have been the better readers; this interpretation is questionable, however.

At present, some schools have the reputation of being "poor" schools because the reading levels of the students are so low in relationship to the reading levels of the average students in the United States. We need to know more precisely the extent to which these students are reading below their potential. It could be that the students in these schools are reading at a level equal to or above their potential. In which case, it may be more appropriate to give the school system a great deal of credit and encouragement. Or, it may be that most of these students are reading far below their expected level. Therefore, it would seem appropriate

to hold the system partially accountable for the fact, and look into ways of changing things for the better.

There needs to be more research into the discrepancy between the actual and expected reading ability of the poorer readers. Certainly, it is unwise, unfair, and unjustifiable to assume that something is wrong with either the students, the teachers, or the school systems when certain students read poorly in comparison to their same age peers. This may be a natural or expected situation that requires no special programs. A better approach would seem to be to isolate those students who are reading poorly in relationship to their own expected reading ability, and give them special attention. Most of what has been said could come under common knowledge for professionals in reading. Yet, it is hoped that saying it again in connection with supporting empirical data will add to its power to influence the policies and behavior of those making practical decisions.

Table 1

Expected Reading Efficiency Scores, $\widehat{RE}$, and Corresponding Expected Grade Level of Reading Ability Scores, $\hat{G}_a$, for Each Rescaled Raven Score, $\hat{I}_a$

$\hat{I}_a$	$\widehat{RE}$	$\hat{G}_a$	$\hat{I}_a$	$\widehat{RE}$	$\hat{G}_a$
0	10-	1.7-	25	38	6.6
1	11	1.9	26	40	6.9
2	12	2.1	28	42	7.2
3	13	2.2	30	44	7.6
4	15	2.6	32	46	7.9
5	16	2.7	34	49	8.4
6	17	2.9	36	51	8.6
7	18	3.1	39	54	8.8
8	19	3.2	42	58	9.1
9	20	3.4	45	61	9.4
10	21	3.6	48	65	9.7
11	22	3.7	52	69	10.0
12	24	4.1	55	73	10.4
13	25	4.2	59	77	10.7
14	26	4.4	63	82	11.1
15	27	4.6	67	86	11.4
16	28	4.7	72	92	11.8
18	30	5.1	76	97	12.2
19	32	5.4	82	103	12.6
20	33	5.6	88	110	13.1
22	35	6.0	94	117	13.7
23	36	6.2	100	124	14.2

References

Bond, G. L. & Tinker, M.A. Reading difficulties: Their diagnosis and correction, 2nd ed. New York, New York: Appleton-Century, Crofts, 1957.

Carver, R. P. Reading tests in 1970 versus 1980: Psychometric versus edumetric tests. The Reading Teacher, 1972, 26, 299-302.

Carver, R. P. Two dimensions of tests: Psychometric and edumetric. American Psychologist, 1974, 29, 512-518.

Carver, R. P. Revised procedures for developing reading-input materials and reading-storage tests. Journal of Reading Behavior, 1975, 2, 155-172.

Carver, R. P. Manual for the National Reading Standards Test. Kansas City, Mo.: Revrac Publications, 1977. (a)

Guilford, J. P. The nature of human intelligence. New York: McGraw-Hill, 1967.

Harris, A. J. How to increase reading ability. 5th ed. New York, New York: David McKay, 1970.

Jensen, A. R. How can we boost IQ and scholastic achievement? Harvard Educational Review, 1969, 39, 1-123.

Spearman, C. Theory of a general factor. British Journal of Psychology, 1946, 36, 117-131.

Thorndike, R. L. Reading as reasoning. Invited address at the annual meeting of the American Psychological Association, Washington, D. C., September 1971.

Vincent, D. V. The linear relationship between age and score of adults in intelligence tests. Occupational Psychology, 1952, 26, 243-249.

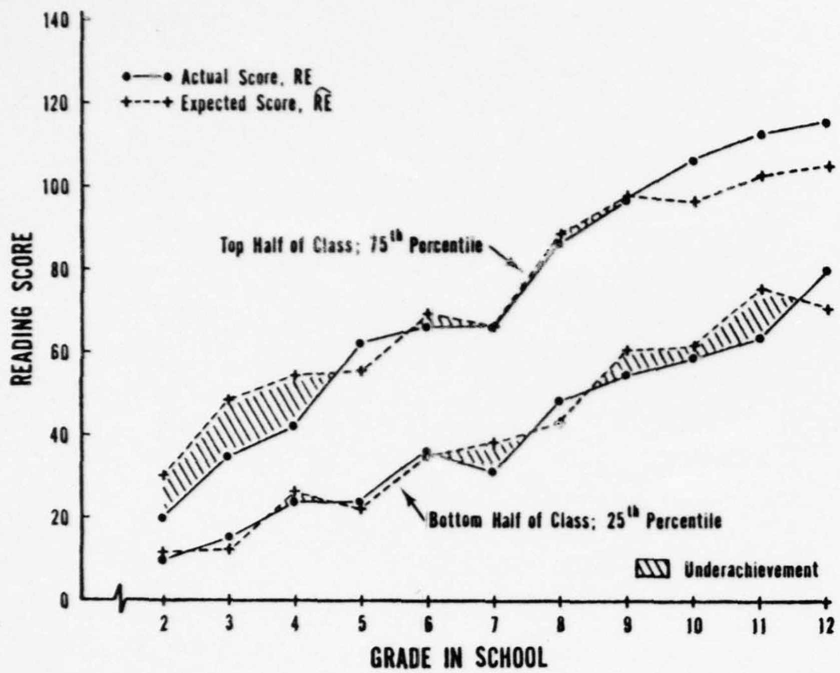


Fig. 1. Actual and expected reading scores as a function of grade in school for the top and bottom halves of each class.

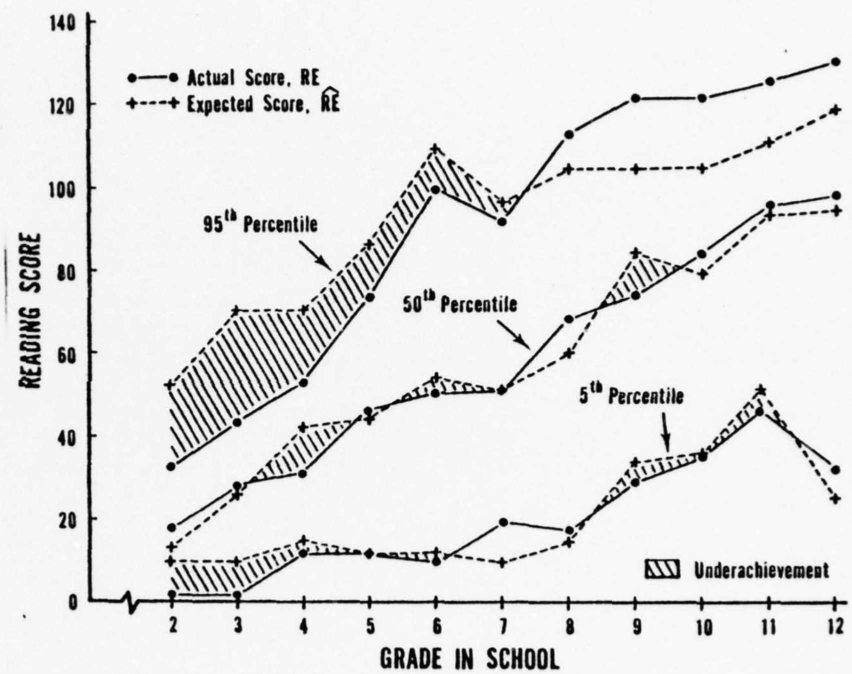


Fig. 2. Actual and expected reading scores as a function of grade in school for the top 10% and the bottom 10% of the class; also included are the actual and expected scores for the class as a whole, the 50th percentile.

DISTRIBUTION LIST

Navy

4 Dr. Marshall J. Farr, Director
Personnel & Training Research Programs
Office of Naval Research (Code 458)
Arlington, VA 22217

1 ONR Branch Office
495 Summer Street
Boston, MA 02210
Attn: Dr. James Lester

1 ONR Branch Office
1030 East Green Street
Pasadena, CA 91101
Attn: Dr. Eugene Gloye

1 ONR Branch Office
536 S. Clark Street
Chicago, IL 60605
Attn: Dr. Charles E. Davis

1 Dr. M. A. Bertin, Scientific Director
Office of Naval Research
Scientific Liaison Group/Tokyo
American Embassy
APO San Francisco 96503

1 Office of Naval Research
Code 200
Arlington, VA 22217

6 Commanding Officer
Naval Research Laboratory
Code 2627
Washington, DC 20390

1 LCDR Charles J. Theisen, Jr., MSC, USN
4024
Naval Air Development Center
Warminster, PA 18974

1 Commanding Officer
U.S. Naval Amphibious School
Coronado, CA 92155

1 Director, Human Resource Management
Naval Amphibious School
Naval Amphibious Base, Little Creek,
Norfolk, VA 23521

1 CDR Paul D. Nelson, MSC, USN
Naval Medical R&D Command (Code 44)
National Naval Medical Center
Bethesda, MD 20014

1 Commanding Officer
Naval Health Research Center
San Diego, CA 92152
Attn: Library

1 Chairman, Leadership & Law Dept.
Div. of Professional Development
U. S. Naval Academy
Annapolis, MD 21402

1 Scientific Advisor to the Chief
of Naval Personnel (Pers Or)
Naval Bureau of Personnel
Room 4410, Arlington Annex
Washington, DC 20370

1 Dr. Jack R. Borsting
Provost & Academic Dean
U. S. Naval Postgraduate School
Monterey, CA 93940

1 Mr. Maurice Callahan
NODAC (Code 2)
Dept. of the Navy
Bldg. 2, Washington Navy Yard
(Anacostia)
Washington, DC 20374

1 Office of Civilian Personnel
Code 263
Washington, DC 20390

1 Superintendent (Code 1424)
Naval Postgraduate School
Monterey, CA 93940

1 Mr. George N. Graine
Naval Sea Systems Command
SEA 047C12
Washington, DC 20362

1 Chief of Naval Technical Training
Naval Air Station Memphis (75)
Millington, TN 38054
Attn. Dr. Norman J. Kerr

1 Principal Civilian Advisor
for Education and Training
Naval Training Command, Code 00A
Pensacola, FL 32508
Attn: Dr. William L. Maloy

1 Dr. Alfred F. Smode, Director
Training Analysis & Evaluation Group
Department of the Navy
Orlando, FL 32813

1 Chief of Naval Education and
Training Support (01A)
Pensacola, FL 32509

1 Capt. H. J. Connery, USN
Navy Medical R&D Command
NNMC, Bethesda, MD 20014

1 Navy Personnel R&D Center
Code 01
San Diego, CA 92152

2 Navy Personnel R&D Center
Code 306
San Diego, CA 92152
Attn: Dr. James McGrath

5 A. A. Sjoholm, Head, Technical Support
Navy Personnel R&D Center
Code 201
San Diego, CA 92152

1 Navy Personnel R&D Center
Code 310
San Diego, CA 92152
Attn: Dr. Martin F. Wiskoff

1 Navy Personnel R&D Center
San Diego, CA 92152
Attn: Library

Navy Personnel R&D Center
San Diego, CA 92152
Attn: Dr. P. Federico

1 Dr. John Ford
Navy Personnel R&D Center
San Diego, CA 92152

1 Dr. Worth Scanland
Chief of Naval Education & Training
NAS, Pensacola, FL 32508

Army

1 Technical Director
U.S. Army Research Institute for the
Behavioral & Social Sciences
5001 Eisenhower Avenue
Alexandria, VA 22333

1 Armed Forces Staff College
Norfolk, VA 23511
Attn: Library

1 Commandant
U. S. Army Infantry School
Fort Benning, GA 31905
Attn: ATSH-I-V-IT

1 Commandant
U. S. Army Institute of Administration
Attn: EA
Fort Benjamin Harrison, IN 46216

1 Dr. Frank J. Harris
U.S. Army Research Institute
5001 Eisenhower Avenue
Alexandria, VA 22333

1 Dr. Ralph Dusek
U.S. Army Research Institute
5001 Eisenhower Avenue
Alexandria, VA 22333

1 Dr. Leon Nawrocki
U.S. Army Research Institute
5001 Eisenhower Avenue
Alexandria, VA 22333

1 Dr. Joseph Ward
U.S. Army Research Institute
5001 Eisenhower Avenue
Alexandria, VA 22333

1 Dr. Ralph Canter
U.S. Army Research Institute
5001 Eisenhower Avenue
Alexandria, VA 22333

1 HQ USAREUE & 7th Army
ODCSOPS
USAREUE Director of GED
APO New York 09403

1 DCDR, USAADMINCEN
Bldg. #1, A310
Attn. AT21-OED Library
Ft. Benjamin Harrison, IN 46216

1 Research Branch
AFMPC/DPMYP
Randolph AFB, TX 78148

1 AFHRL/AS (Dr. G. A. Eckstrand)
Wright-Patterson AFB
Ohio 45433

1 Dr. Ross L. Morgan (AFHRL/ASR)
Wright-Patterson AFB
Ohio 45433

1 Dr. Marty Rockway (AFHRL/TT)
Lowry AFB
Colorado 80230

1 Instructional Technology Branch
AFHRL
Lowry AFB, CO 80230

1 Dr. Alfred R. Fregly
AFOSR/NL, Building 410
Bolling AFB, DC 20332

1 Dr. Sylvia R. Mayer (MCIT)
HQ Electronic Systems Division
LG Hanscom Field
Bedford, MA 01730

1 Air Force Human Resources Lab
AFHRL/PED
Brooks AFB, TX 78235

1 Major Wayne S. Sellman
Chief, Personnel Testing
AFMPC/DPMYO
Randolph AFB, TX 78148

1 Air University Library
AUL/LSE 76-443
Maxwell AFB, AL 36112

Marine Corps

1 Director, Office of Manpower
Utilization
HQ, Marine Corps (Code MPU)
BCB, Building 2009
Quantico, VA 22134

1 Dr. A. L. Slafkosky
Scientific Advisor (Code RD-1)
HQ, U.S. Marine Corps
Washington, DC 20380

1 AC/S, Education Programs
Education Center, MCDEC
Quantico, VA 22134

Coast Guard

1 Mr. Joseph J. Cowan, Chief
Psychological Research Branch
U.S. Coast Guard Headquarters
Washington, DC 20590

Other DoD

1 Dr. Harold F. O'Neil, Jr.
Advanced Research Projects Agency
Cybernetics Technology, Room 623
1400 Wilson Blvd.
Arlington, VA 22209

1 Mr. Frederick W. Suffa
Chief, Recruiting and Retention Evaluation
Office of the Assistant Secretary of
Defense, M&RA
Room 3D970, Pentagon
Washington, DC 20301

12 Defense Documentation Center
Cameron Station, Bldg. 5
Alexandria, VA 22314
Attn: TC

1 Military Assistant for Human Resources
Office of the Director of Defense
Research & Engineering
Room 3D129, The Pentagon
Washington, DC 20301

1 Director, Management Information
Systems Office
OSD, M&RA
Room 3B917, the Pentagon
Washington, DC 20301

PRECEDING PAGE BLANK-NOT FILMED

Other Government

1 Dr. William Gorham, Director
Personnel R&D Center
U.S. Civil Service Commission
1900 E Street NW
Washington, DC 20415

1 Dr. Andrew R. Molnar
Science Education Dev. & Res.
National Science Foundation
Washington, DC 20550

1 Dr. Marshall S. Smith
Associate Director
NIE/OPEPA
National Institute of Education
Washington, DC 20208

1 Dr. Joseph L. Young, Director
Memory & Cognitive Processes
National Science Foundation
Washington, DC 20550

1 Dr. H. Wallace Sinaiko, Director
Manpower Research & Advisory Services
Smithsonian Institution
801 N. Pitt Street
Alexandria, VA 22314

Miscellaneous

1 Dr. John R. Anderson
Dept. of Psychology
Yale University
New Haven, CT 06520

1 Dr. Scarvia B. Anderson
Educational Testing Service
Suite 1040
3445 Peachtree Road NE
Atlanta, GA 30326

1 Mr. Samuel Ball
Educational Testing Service
Princeton, NJ 08540

1 Dr. Gerald V. Barrett
University of Akron
Dept. of Psychology
Akron, OH 44325

1 Dr. John Seeley Brown
Bolt Beranek and Newman, Inc.
50 Moulton Street
Cambridge, MA 02138

1 Century Research Corporation
4113 Lee Highway
Arlington, VA 22207

1 Dr. Kenneth E. Clark
College of Arts & Sciences
University of Rochester
River Campus Station
Rochester, NY 14627

1 Dr. Norman Cliff
Dept. of Psychology
University of Southern California
University Park
Los Angeles, CA 90007

1 Dr. Allan M. Collins
Bolt Beranek and Newman, Inc.
50 Moulton Street
Cambridge, MA 02138

1 Dr. Rene V. Dawis
Dept. of Psychology
University of Minnesota
Minneapolis, MN 55455

1 Dr. Ruth Day
Dept. of Psychology
Yale University
Box 11A, Yale Station
New Haven, CT 06520

1 Dr. John D. Carroll
Psychometric Lab
Davie Hall 013A
University of North Carolina
Chapel Hill, NC 27514

1 Dr. Marvin D. Dunnette
Dept. of Psychology
University of Minnesota
Minneapolis, MN 55455

1 ERIC Facility-Acquisitions
4833 Rugby Avenue
Bethesda, MD 20014

1 Major I. N. Evonic
Canadian Forces Personnel
Applied Research Unit
1107 Avenue Road
Toronto, Ontario, CANADA

1 Dr. Richard L. Ferguson
The American College Testing Program
P. O. Box 168
Iowa City, IA 52240

1 Dr. Victor Fields
Dept. of Psychology
Montgomery College
Rockville, MD 20850

1 Dr. Edwin A. Fleishman
Advanced Research Resources Organization
8555 Sixteenth Street
Silver Spring, MD 20910

1 Dr. John R. Frederiksen
Bolt Beranek & Newman, Inc.
50 Moulton Street
Cambridge, MA 02138

1 Dr. Vernon S. Gerlach
College of Education
146 Payne Bldg. B
Arizona State University
Tempe, AZ 85281

1 Dr. Robert Glaser, Co-Director
University of Pittsburgh
3939 O'Hara Street
Pittsburgh, PA 15213

1 CDR Mercer
CNET Liaison Officer
AFHRL/Flying Training Div.
Williams AFB, AZ 85224

1 HumRRO/Western Division
27857 Berwick Drive
Carmel, CA 93921
Attn: Library

1 HumRRO/Western Division
27857 Berwick Drive
Carmel, CA 93921
Attn: Dr. Robert Vineberg

1 Dr. Lawrence B. Johnson
Lawrence Johnson & Associates, Inc.
Suite 502
2001 S Street NW
Washington, DC 20009

1 Dr. Arnold F. Kanarick
Honeywell, Inc.
2600 Ridgeway Pkwy.
Minneapolis, MN 55413

1 Dr. Roger A. Kaufman
203 Dodd Hall
Florida State University
Tallahassee, FL 32306

1 Dr. Steven W. Keele
Dept. of Psychology
University of Oregon
Eugene, OR 97403

1 Dr. David Klahr
Dept. of Psychology
Carnegie-Mellon University
Pittsburgh, PA 15213

1 Mr. Brian McNally
Educational Testing Service
Princeton, NJ 08540

1 Dr. Robert R. Mackie
Human Factors Research, Inc.
6780 Corton Drive
Santa Barbara Research Park
Goleta, CA 93017

1 Mr. Edmond Marks
304 Grange Bldg.
Pennsylvania State University
University Park, PA 16802

1 Dr. Leo Munday
Houghton Mifflin Co.
P. O. Box 1970
Iowa City, IA 52240

1 Dr. Donald A. Norman
Dept. of Psychology C-009
University of California, San Diego
La Jolla, CA 92093

1 Mr. Luigi Petrullo
2431 N. Edgewood Street
Arlington, VA 22207

1 Dr. Steven M. Pine
N 660 Elliott Hall
University of Minnesota
75 East River Road
Minneapolis, MN 55455

1 Dr. Lyman W. Porter, Dean
Graduate School of Administration
University of California
Irvine, CA 92717

1 Dr. Diane M. Ramsey-Klee
R-K Research & System Design
3947 Ridgemont Drive
Malibu, CA 90265

1 R.Dir. M. Rauch
P II 4
Bundesministerium der Verteidigung
Postfach 161
53 Bonn 1, GERMANY

1 Dr. Joseph W. Rigney
University of So. California
Behavioral Technology Laboratories
3717 South Grand
Los Angeles, CA 90007

1 Dr. Andrew M. Rose
American Institutes for Research
1055 Thomas Jefferson St. NW
Washington, DC 20007

1 Dr. Leonard L. Rosenbaum, Chairman
Dept. of Psychology
Montgomery College
Rockville, MD 20850

1 Dr. Mark D. Reckase
Educational Psychology Dept.
University of Missouri-Columbia
12 Hill Hall
Columbia, MO 65201

1 Dr. Robert J. Seidel
Instructional Technology Group,
HumRRO
300 N. Washington St.
Alexandria, VA 22314

1 Dr. Richard Snow
Stanford University
School of Education
Stanford, CA 94305

1 Mr. Dennis J. Sullivan
c/o Canyon Research Group, Inc.
32107 Lindero Canyon Road
Westlake Village, CA 91360

1 Dr. Benton J. Underwood
Dept. of Psychology
Northwestern University
Evanston, IL 60201

1 Dr. Carl R. Vest
Battelle Memorial Institute
Washington Operations
2030 M Street NW
Washington, DC 20036

1 Dr. David J. Weiss
Dept. of Psychology
N660 Elliott Hall
University of Minnesota
Minneapolis, MN 55455

1 Dr. Keith Wescourt
Dept. of Psychology
Stanford University
Stanford, CA 94305

1 Dr. Earl Hunt
Dept. of Psychology
University of Washington
Seattle, WA 98105

1 Dr. Thomas G. Sticht
Assoc. Director, Basic Skills
National Institute of Education
1200 19th Street NW
Washington, DC 20208

1 Prof. Fumiko Samejima
Dept. of Psychology
Austin Peay Hall 304C
University of Tennessee
Knoxville, TN 37916

1 Dr. Frank Pratzner
The Center for Vocational Education
Ohio State University
1960 Kenny Road
Columbus, Ohio [illegible]

1 Dr. Meredith Crawford
5605 Montgomery Street
Chevy Chase, MD 20015