R E P O R T R E S U M E S

ED 020 098 RE 001 274

FOUR INSTRUCTIONAL APPROACHES TO BEGINNING READING--THREE YEARS LATER.

BY- HAYES, ROBERT B. WUEST, RICHARD C.

PUB DATE APR 68

EDRS PRICE MF-$0.25 HC-$0.68 15P.

DESCRIPTORS- *READING RESEARCH, *READING INSTRUCTION, ELEMENTARY SCHOOL STUDENTS, READING PROGRAMS, READING TESTS, READING SKILLS, DIRECTED READING ACTIVITY, *BEGINNING READING, EARLY READING, DEVELOPMENTAL READING, PHONICS, BASIC READING, INITIAL TEACHING ALPHABET, *PRIMARY GRADES,

FOUR APPROACHES TO BEGINNING READING WERE EVALUATED IN A 3-YEAR STUDY OF ELEMENTARY PUBLIC SCHOOL STUDENTS IN NEW CASTLE, PENNSYLVANIA. THE FOUR APPROACHES USED WERE (1) A BASAL READER PROGRAM PUBLISHED BY SCOTT, FORESMAN AND CO. (1962), (2) A PHONIC PROGRAM UTILIZING CORRELATED FILMSTRIPS PUBLISHED BY J.B. LIPPINCOTT CO. (1963), (3) A COMBINATION USING SCOTT, FORESMAN'S MATERIALS SUPPLEMENTED WITH PHONIC BOOKLETS (PHONICS AND WORD POWER) PUBLISHED BY AMERICAN EDUCATION PUBLICATIONS, INC. (1964), AND (4) A LANGUAGE ARTS APPROACH USING THE INITIAL TEACHING ALPHABET (1963). AFTER THE PUPILS IN THIS GROUP MADE THE TRANSITION TO TRADITIONAL ORTHOGRAPHY, THE TREASURY OF LITERATURE SERIES PUBLISHED BY CHARLES E. MERRILL (1960) WAS USED. IN ADDITION, WIDE INDEPENDENT READING WAS ENCOURAGED IN ALL FOUR GROUPS. THE STANFORD ACHIEVEMENT TEST AND THE GILMORE ORAL READING TEST WERE USED AS PRIMARY MEANS OF EVALUATION FOR THE STUDY. RESULTS ON THESE TESTS SHOWED THAT IN GENERAL THE LIPPINCOTT AND I/T/A-LIPPINCOTT PROGRAM MIGHT BE WORTHY OF ATTENTION AND FUTURE STUDY. HOWEVER, THE RESULTS DO NOT SUGGEST THAT ANY OF THE FOUR APPROACHES WAS CONSISTENTLY BETTER THAN THE OTHERS. THIS PAPER WAS PRESENTED AT THE INTERNATIONAL READING ASSOCIATION CONFERENCE (BOSTON, APRIL 24-27, 1968). (BS)

IRA Boston, 1968

Robert B. Hayes

Director of Research Administration
and Coordination
Department of Public Instruction
Harrisburg, Pennsylvania

Richard C. Wuest

Coordinator of Reading Clinic Services
State University College at Oswego
Oswego, New York

Four Instructional Approaches To Beginning Reading -- Three Years Later

Session: Research Reports - Follow-Up First Grade Studies

Problem

Beginning in September 1964 and continuing until June 1967, a longitudinal study was conducted in the public schools of New Castle, Pennsylvania to determine which of four approaches to beginning reading instruction was the most effective. In addition, a modified replication of the original study was begun in September 1965 as an additional check upon the validity and reliability of obtained results. This replicative study was also concluded at the close of the school year last June.

Method

The independent treatment variables in both studies were: (1) a basal reader program published by Scott, Foresman and Company, 1962

edition; (2) a phonic program utilizing correlated filmstrips and published by J. B. Lippincott Company, 1963 edition; (3) a combination program which used the materials of the Scott, Foresman Company (No. 1 above) supplemented with phonic booklets (Phonics and Word Power) published by American Education Publications, Inc., 1964 edition; and (4) a language arts approach using the initial teaching alphabet as a medium, represented by the materials of i/t/a Publications, Inc., 1963 edition. The Treasury of Literature Series of Charles E. Merrill Books, Inc., 1960 edition was used after i/t/a pupils made transition to traditional orthography. Teachers were restricted to using only those methods and materials recommended by book company consultants for instructional purposes, but wide independent reading was encouraged.

The following dependent variables were included during each year of the study: (1) a group, standardized silent reading achievement test (Stanford Achievement Test); (2) a reading attitude inventory (San Diego County Inventory of Reading Attitude); and (3) a record of the number of books read independently. In addition, randomly selected samples of both populations were given certain individual tests of oral reading achievement including: (1) the Gates Word Pronunciation Test; (2) the Fry Phonetically Regular Words Oral Reading Test; and (3) the Gilmore Oral Reading Test. In January and May of first grade the Primary I Battery of the Stanford was administered, while the Primary II Battery was used in January and May of second and third grades, and the Intermediate I Battery was administered in June of the third grade.

The population of the study was randomly assigned, by attendance areas, to the required number of classrooms and treatment groups. The original study first included five classes for each of the treatments, but the illness of one Scott, Foresman teacher during first grade resulted in the loss of her class from the study. Therefore, nineteen classes and 365 pupils were included in the comparisons drawn at the end of Grade I; 302 pupils remained at the end of Grade II; and third year comparisons were made on the 262 students who remained at the end of Grade III.

In the replicative study, only three classes were selected for each of the treatment groups. Comparisons were made on 248 first grade students and 213 second grade students.

Consultant services were provided to the teachers of the original study by the participating book companies. This was done to assist the teachers in following appropriate procedures, and to help them to understand the philosophies of the companies whose materials were being used. The consultants conducted classroom observations followed by in-service workshop meetings for the teachers in the original study, but since teachers who participated in the replication were almost always those who were in the original study, classroom observations by the consultants and in-service work was largely eliminated in an attempt to control Hawthorne effects.

Frequent random, unannounced classroom observations by administrative personnel were employed to determine the extent of teachers'

adherence to prescribed procedures and to evaluate teaching effectiveness. During these observation periods, each supervisor independently rated the teachers on the Hayes Teacher Rating Scale. Twenty classroom visits were made during the first year of the original study, and twelve visitations were made to each classroom of both grades during the second and third years.

All teachers were also required to submit logs to the field director as another method safeguard. Teachers in the original study kept logs during alternate weeks on which they summarized the objectives for each lesson, the skills taught, the materials used, the grouping procedures followed, and the time spent teaching reading for each day. Since almost all teachers in the replication had kept and submitted logs during alternate weeks in previous years, they were only required to record a summary of the materials used and the grouping procedures followed at the end of each month. This variation in requirements was followed as a means of further reducing Hawthorne effects in the replicative study.

The local school district required that reading be taught for 560 minutes per week during the first grade, 530 minutes per week during second grade, and 415 minutes per week in third grade.

Statistical Analysis

Statistical analysis consisted of correlation coefficients, a 4 x 3 factorial analysis of variance and covariance (where appropriate). In this analysis, factor A consisted of four methods of teaching reading,

while factor B represented three levels of intelligence (high, average, and low). In the third year of the study the preceding analysis involved random casting out of cases to produce an equal number of cases per cell. This resulted in 15 cases per IQ level, 45 per treatment and a total N of 180 in Grade II and also in Grade III. The Stanford paragraph meaning scores were also analyzed for all students by an unweighted means analysis with very similar results to the analysis for just 180 pupils.

For the analysis of variance involving 180 cases per grade, a Tukey (a) multiple range test was employed to determine which differences between means were contributing to significant F ratios. When analysis of covariance produced significant F ratios, Winer's multiple F test was used to compare differences between each appropriate pair of means.

Bond and Tinker reading expectancy scores were compared to grade equivalent scores for Word Reading, Word Study Skills and Paragraph Meaning of the Stanford Achievement Test.

The analysis of variance, covariance and correlation matrices were performed at the Computation Center of The Pennsylvania State University, University Park, Pennsylvania, in the final year of the study, while in the first two years the data was analyzed by the University of Minnesota Computer.

Results

While only a summary of some of the major findings for the original study are reported in this paper, the replicative study results

largely confirm the findings of the original study.

The mean intelligence quotients for the third grade treatment groups were: Scott, Foresman - 98.49; Lippincott - 98.58; Phonics and Word Power - 96.98; and i/t/a-Merrill - 97.96. The mean IQ's, by levels, of the various treatment groups were:

	SF	Lipp	PWP	i/t/a-Merr
High IQ	112.40	114.07	108.87	112.66
Average IQ	99.67	98.93	98.40	97.07
Low IQ	83.40	82.73	83.67	84.13

Third grade average teacher effectiveness ratings were also very similar: 15.67 for Scott, Foresman; 15.40 for Lippincott; 15.18 for Phonics and Word Power; and 14.40 for i/t/a-Merrill.

The grade equivalent means on the paragraph meaning subtest of the Stanford Achievement Test during the three years of the original study are presented in the following tables. Whenever it was necessary, the scores were adjusted statistically for factors such as intelligence and teacher effectiveness ratings, and original comparisons were based upon raw scores. Grade equivalent scores are reported as a convenience to the reader.

Robert B. Hayes
Richard C. Wuest

TABLE 1

PARAGRAPH MEANING BY TREATMENTS

	SF	PWP	Lipp	i/t/a-Merr
Grade I - January 1965	1.4	1.6	1.6	1.6
Grade I - April 1965	1.7	1.8	1.8	1.8
Grade II - January 1966	2.6	2.5	2.9	2.8
Grade II - May 1966	2.9	3.2	3.1	3.1
Grade III - January 1967	3.4	3.7	3.8	3.8
Grade III - June 1967	4.3	4.4	4.9	4.6

For the above, significant differences occurred as follows: (1) in January of Grade II when Lippincott was compared to Phonics and Word Power, and (2) in June of Grade III when Lippincott was compared with Phonics and Word Power and also with Scott, Foresman.

The paragraph meaning grade equivalent means on the Stanford Achievement Test for the high, average, and low IQ levels were as follows:

TABLE 2

PARAGRAPH MEANING BY TREATMENTS

(HIGH IQ LEVEL)

	SF	PWP	Lipp	i/t/a-Merr
Grade I - April 1965	2.0	1.9	2.4	2.4
Grade II - January 1966	2.9	2.9	3.4	3.3
Grade II - May 1966	3.4	3.6	3.8	3.9
Grade III - January 1967	3.8	3.9	4.3	4.0
Grade III - June 1967	4.8	4.7	6.0	4.9

Robert B. Hayes
Richard C. Wuest

For the above, the significant differences were: (1) in April of Grade I and in January of Grade II, i/t/a-Merrill and Lippincott compared to Scott, Foresman and Phonics and Word Power; (2) in May of Grade II, i/t/a-Merrill and Lippincott compared to Scott, Foresman, and (3) in June of Grade III, Lippincott versus Phonics and Word Power.

TABLE 3

PARAGRAPH MEANING BY TREATMENTS

(AVERAGE IQ LEVEL)

	SF	PWP	Lipp	i/t/a-Merr
Grade I - April 1965	1.8	1.7	1.9	1.9
Grade II - January 1966	2.7	2.5	2.9	2.8
Grade II - May 1966	3.0	3.1	3.1	3.1
Grade III - January 1967	3.5	3.7	4.0	4.0
Grade III - June 1967	4.7	4.6	4.9	4.8

For the above, the significant differences were: (1) in April of Grade I, Lippincott and i/t/a-Merrill compared to Phonics and Word Power, and (2) in January of Grade II, Lippincott versus Phonics and Word Power.

Robert B. Hayes
Richard C. Wuest

TABLE 4

PARAGRAPH MEANING BY TREATMENTS

(LOW IQ LEVEL)

	SF	PWP	Lipp	i/t/a-Merr
Grade I - April 1965	1.6	1.6	1.6	1.6
Grade II - January 1966	2.4	2.1	2.4	2.1
Grade II - May 1966	2.9	2.6	2.6	2.6
Grade III - January 1967	3.0	3.3	3.3	3.3
Grade III - June 1967	3.7	3.9	4.2	3.9

The above differences were not significant.

The results attained on the Gates Word Pronunciation Test significantly favored Lippincott and i/t/a-Merrill over Scott, Foresman and Phonics and Word Power at the end of first and second grades, but there were no significant differences on this variable at the end of Grade III.

On reading accuracy, as measured by the Gilmore Oral Reading Test, there was only one significant difference found at the end of Grade I. This involved children in the high IQ third where Lippincott and i/t/a-Merrill were ahead of both Scott, Foresman and Phonics and Word Power. In May of Grade II for the average IQ third, i/t/a-Merrill was significantly ahead of Scott, Foresman. By the end of Grade III, the following significant differences were found: (1) for the entire subsample, i/t/a-

Merrill over Scott, Foresman, and (2) for the high IQ third, Lippincott and i/t/a-Merrill were ahead of Scott, Foresman.

Reading comprehension results, as measured by the Gilmore produced only one ...ificant difference at the end of first grade and this involved children in the average IQ third where Scott, Foresman led Lippincott. There were no significant differences in May of Grade II, but at the end of third grade, the following differences proved to be significant: (1) for the total subsample, i/t/a-Merrill was favored over Lippincott and Scott, Foresman; (2) for the high IQ third, Lippincott and i/t/a-Merrill were ahead of Scott, Foresman and Phonics and Word Power; and (3) for the low IQ third, i/t/a-Merrill led Lippincott.

The only significant differences in rate of reading on the Gilmore occurred at the end of Grade I as follows: (1) for the entire subsample, i/t/a-Merrill was higher than Lippincott and Phonics and Word Power; (2) for the high IQ third, i/t/a led Phonics and Word Power; and (3) for the average IQ third, i/t/a-Merrill and Scott, Foresman were ahead of Lippincott.

At the end of Grade I, the Phonics and Word Power group received significantly higher ratings than each of the other three treatment groups on the San Diego County Inventory of Reading Attitude, while in April of Grade II, Scott, Foresman was rated significantly lower than the others. Third year results indicated no significant differences among the groups in reading attitude as measured by the San Diego.

When attitude toward reading was measured by comparing the number

of books read independently in a typical month by each treatment group, the following significant differences were discovered: (1) in Grade I, Scott, Foresman led each of the other three treatment groups and Lippincott read more than either Phonics and Word Power or i/t/a-Merrill; (2) in Grade II, i/t/a-Merrill was behind each of the other treatment groups; and (3) in Grade III, Lippincott and Scott, Foresman read more than either of the other two treatment groups.

Discussion and Conclusions

Two instruments of evaluation were used in this study to investigate differences in the reading comprehension of the four treatment groups. These were: (1) the Stanford Achievement Test and (2) the Gilmore Oral Reading Test. A close analysis of the results reveals some interesting contrasts. For example, when the comprehension of the entire population of the study was compared by treatment groups on the basis of Stanford results, the Lippincott group significantly led Phonics and Word Power at the end of Grades II and III and was also ahead of the Scott, Foresman group at the completion of third grade. However, comprehension results based upon the Gilmore for the entire subsample indicated that the only significant difference which existed occurred at the end of Grade III when the i/t/a-Merrill group significantly led both Lippincott and Scott, Foresman. Performance on these particular tests of comprehension probably require different complexes of skills, but these results indicate that an i/t/a-Lippincott program might be worthy of attention and future study.

Robert B. Hayes
Richard C. Wuest

A closer analysis of the significant mean differences on the comprehension tests by treatments and IQ thirds reveals some other interesting relationships. There were many more significant differences attained through the years on the Stanford than there were on the Gilmore. Comprehension differences on the Stanford generally favored both i/t/a-Merrill and Lippincott during the first two years of the study (especially for the high IQ third), but only one significant comprehension difference was found on the Gilmore during first and second grades (at the end of Grade I, for the average IQ third, Scott, Foresman was significantly ahead of Lippincott).

Greater agreement seems to exist among the comprehension results of both tests at the end of Grade III where for the high IQ third on both tests, Lippincott scored significantly higher than Phonics and Word Power, and on the Gilmore, Lippincott also led Scott, Foresman. Gilmore results indicated that i/t/a-Merrill also significantly led Phonics and Word Power and Scott, Foresman. For the average IQ third, there were no significant differences on either test at the end of Grade III. For the low IQ third at the end of the third grade, the Stanford revealed no significant differences, but Gilmore results placed i/t/a-Merrill significantly ahead of Lippincott.

In attempting to determine the effects of each instructional approach upon the ability to read words orally, the results attained on two tests have been reported: the Gates Word Pronunciation Test measures the ability to read lists of isolated words, and the Gilmore

Oral Reading Test provides a measure of accuracy of orally reading words in contextual settings.

Lippincott and i/t/a-Merrill achieved significantly greater word recognition scores on the Gates than Scott, Foresman or Phonics and Word Power in Grades I and II, but by the end of Grade III, there were no significant differences among the groups for this variable. In contrast, the Gilmore revealed no significant differences until the end of Grade III when i/t/a-Merrill significantly led Scott, Foresman. Since the overall comprehension results on the Gilmore were not significantly different until the end of Grade III, reading programs with a heavy decoding emphasis apparently gave children greater power in recognizing isolated words in lists, but this advantage may not be readily transferred in Grades I and II to deriving understanding when reading words in context in oral reading situations.

A comparison of the reading accuracy on the Gilmore by treatment groups and IQ thirds revealed that for the high IQ third, Lippincott and i/t/a-Merrill were generally favored over the other treatment groups, while for the average IQ third, the only significant difference favored i/t/a-Merrill over Scott, Foresman at the end of Grade II, and there were no significant differences for the low IQ third.

The only significant differences which were found in the rate of oral reading as measured by the Gilmore occurred at the end of Grade I and generally favored the i/t/a-Merrill group.

In Grade I the Phonics and Word Power group scored highest on the

San Diego County Inventory of Reading Attitude, but read comparatively few books independently, while Scott, Foresman children read significantly more books than the others. Grade II results on the reading attitude inventory placed Scott, Foresman significantly behind the other groups, but by the end of Grade III, there were no significant differences among the groups. In all three grades i/t/a-Merrill pupils lagged significantly behind on the number of books read other than regular textbooks.

Although not reported elsewhere in this paper, it is important to note that dutting 1964-1965, twelve percent of the Lippincott pupils were retained in Grade I compared to three percent of the i/t/a pupils, six percent of the Scott, Foresman pupils, and six percent of the Phonics and Word Power pupils. In the second year of the study, 1965-1966, there were almost eight percent of the Lippincott children who were retained in second grade compared to almost five percent i/t/a-Merrill pupils, almost two percent Scott, Foresman pupils, and almost five percent Phonics and Word Power pupils.

Implications

This study indicated that both methods and materials can make a difference in teaching reading. In general, the Lippincott and i/t/a-Merrill groups seemed to make the best progress as measured by the evaluation instruments which were used in this study. However, the results do not suggest that any of the approaches which were investigated are consistently better than others.

Robert B. Hayes
Richard C. Wuest

Each of the four approaches to beginning reading instruction included in this study were used under rather ideal conditions. The in-service education provided to the teachers was generally excellent. Teachers received more supervision than is normally available. All of the most recent materials offered by the involved companies were provided. Therefore, it cannot be assumed that any one of the approaches, without the conditions of this study, would produce the same results.

Finally, while silent and oral reading were evaluated in this study and various relationships were determined, other relationships among language, thinking, and beginning reading instruction were not investigated. Reading is one aspect of the total language process, and is therefore closely related to other language abilities, both affecting and being affected by them. Through the use of language, thinking is facilitated and ideas are communicated through abstract symbols. Language could not exist without thought and thinking would be severely limited without language. This inseparable unification of language and thought processes suggests the desirability of future investigations of beginning reading instruction to include the refinement of existing evaluative techniques and the development of new measuring devices which could be used to assess important relationships among other language abilities, thinking, and various approaches to beginning reading instruction.