

Learning to Read: The Great Debate.

Part 2,

The Evidence: Research on Beginning Reading Instruction

FOUR

Experimental Evidence

on Approaches to

Beginning Reading

Pgs. 99 - 139

WHEN I SET ABOUT examining the experimental research, I found myself facing some important problems: What research should be considered? How was it to be found? How should it be read and analyzed? How could the results of all these studies be synthesized into some valid findings and conclusions? And so forth.

These may appear to be naïve concerns, especially when one is dealing with a body of research that has been reviewed many times before. However, reviews of this same research had led to very different conclusions and recommendations. Clearly, the way in which to review it was a crucial point.

SELECTING AND ORGANIZING THE EXPERIMENTS

I selected the experimental studies from references in all the relevant sources known to me. I do not claim to have included all experimental studies on the big question of a code or meaning emphasis, but I believe that I sought out, with equal persistence, references cited by both the critics and the defenders of currently prevailing meaning-emphasis pro-

grams. The bibliography at the end of the book lists all the studies considered.

The studies chosen were then sorted according to the questions they seemed to be asking. I narrowed them down to four groups concentrated on the following four questions:

1. Which achieves better results—a look-say or a phonic method?
2. Which achieves better results—a method that teaches more phonics or one that teaches less?
3. What do experimental comparisons involving a linguistic approach tell us about the values of such an approach?
4. What do experimental comparisons involving the use of modified alphabets tell us about the benefits of these innovations?

I also asked a fifth question: Do the experimental studies indicate that the effectiveness of the various methods depends on characteristics of the individual child being taught—his mental ability and his socioeconomic background? In other words, is one method better or worse for the bright or the dull, for the rich or the poor?

EXTRACTING THE "TRUTH" FROM EACH STUDY

Since previous summaries based on practically the same body of experimental research had arrived at conflicting conclusions, I knew before starting that a major problem was how to read the research. Reading abstracts of the studies and their authors' conclusions would not suffice. Instead, I would have to look at each study carefully and to ask such questions as why it was made; how the author defined "reading"; what methods and materials were being compared; what the size, age, grade level, and other characteristics of the population studied were; how the author defined reading "success"; what care was taken to assure comparability of the groups studied; and so on. To this end my associates and I listed all the important conditions that could influence the outcome of a study, and we checked each study against the list (see Appendix B, Schedule I).

As I had suspected beforehand, practically none of the studies specified all these conditions. Most did not indicate how the experimental and control groups were selected, how much time was allotted to various aspects of reading, how the teachers were selected, whether the quality of the teaching was comparable in both groups, or even whether the

teachers followed the methods under study.¹ Even more important, most studies did not specify clearly what a "method" involved, but instead merely assigned labels (e.g., "phonics"), expecting the reader to understand what was meant.

Many of the early studies did not use standard measures of outcomes or statistical tests of significance to determine whether the various results obtained could have been attributable to chance differences. At the same time, some of the first studies, with their small populations, "homemade" tests, and simple statistical techniques, had many strengths that the later, more statistically sophisticated ones lacked, and I included them for that reason. The studies of Buswell (1922), Gates (1927), Gates and Russell (1938), Winch (1925), and Valentine (1913), for example, were more thoroughly grounded in theory (whether of the learning or of the reading process), used more imaginative instruments to test the different components of reading, and were more analytic in presenting test results. Also, they tended to "look at the learners," describing in considerable detail how the children under study approached words, what errors they made, their attitudes toward reading, how their teachers reacted, etc.²

This kind of analysis is virtually nonexistent in the recent comparisons, which often present results only in terms of tests of significance on various subtests of standard instruments. Seldom do the investigators try to explain what in fact these various subtests are measuring.³ Notable exceptions are the studies by MacKinnon (1959) and Daniels and Diack (1956 and 1960). The latter devised instruments to measure different outcomes, and MacKinnon did a very rare and worthwhile thing—he observed the day-to-day reading performance of children on different kinds of reading material, in addition to testing at the end of the experiment.

Although most studies were unsatisfactory in some respects, I assumed that all the authors were honest researchers searching for honest answers, and I looked for the grains of underlying truth to be found in

¹ Indeed, in one of the recent USOE first-grade studies referred to earlier and summarized briefly at the end of this chapter, we found that there was a considerable variation in the way the teachers followed a given method and in the results they achieved with it. We found little relationship between what they said they did and what they actually did during reading lessons. But what they actually did with the method made a difference in pupil achievement (Chall and Feldmann, 1966).

² See especially Gates (1928).

³ See in this connection the most recent of these comparisons—the cooperative first-grade studies supported by the USOE and reported in *The Reading Teacher*, May, 1966.

each study. Had I considered only studies that fulfilled all necessary experimental conditions, I would have been left with just a handful—if that many.

In analyzing the studies I gave priority to the authors' findings rather than their conclusions, and for a very good reason: Many authors drew conclusions that seemed to go counter to their own findings. Every researcher runs the risk of generalizing his findings beyond his data. To counter this tendency, it is necessary to consider the author's conclusions in terms of both his findings and his assumptions. For example, Mosher and Newhall (1930) concluded that there is little advantage in a phonic over a look-say method, although results in eight out of the ten tests they gave the children favored phonics slightly. They justified their conclusion on the grounds that the two largest differences favored the look-say approach. Because Mosher's earlier study (1928) had also concluded that a look-say was better than a phonic method, although no comparison was made of the look-say experimental group with a control group taught by a phonic method, I felt that the investigators were perhaps too favorably disposed toward their own method and would tend to pull any "uncertainties" of their findings in the direction of their preference.

CLASSIFYING METHODS AND MATERIALS STUDIED

When a researcher says he is testing a phonic method, what does he mean?

As I have mentioned, few of the studies specified clearly what the methods being studied entailed. Often the labels and titles were misleading. Thus, Currier's study "Phonics and No Phonics" (1923) really compares varying amounts and kinds of phonics. Many investigators speak of *phonetics* (the scientific study of speech sounds) when they mean some form of *phonics*.

To account for this problem I set up my own classification system. I called a method "look-say" if the author stated that it taught no phonics at all and emphasized visual recognition of the whole word, "getting the thought," and reading whole sentences.

I classified as "systematic-phonics" programs those which taught phonics early and systematically—usually, but not always, before sight (whole) words. Such programs usually taught phonics separately from connected reading. The approach was often synthetic rather than analytic (i.e., the children were taught the separate letter-sound correspondences and were given practice in blending these sounds). Not all the programs classified as "systematic phonics" programs met all these criteria. However, they all put an *earlier, heavier, and more direct* emphasis on teach-

ing the sound values of the letters than other programs with which they were compared.

I classified as "intrinsic-phonics" programs those which stressed sight or thought reading, introduced phonics later, and taught a more moderate amount of it—all *intrinsic* to meaningful reading, which was the supreme consideration. Children following these programs usually learned the sound values of letters through a process of analyzing known sight words. Other means of identifying words, e.g., context and picture clues, received greater stress than word analysis. Generally no separate period was set aside for phonics practice. There was some variation, however, on this last point and on the extent to which the phonics taught followed a preordered "system" or was based solely on the pupils' needs as determined by the teacher.

For linguistic and modified alphabet treatments I accepted the authors' classifications. However, as we shall see later, some programs that investigators labeled "linguistic" or "modified linguistic" can just as readily be classified as "systematic phonics."

The modified alphabet studies were easily classified.

Thus, we can place the systematic-phonics treatments at the far end of the code side of our code-meaning continuum; near them we can also place most of the linguistic approaches and the alphabet reforms. At the far end of the meaning side we can place the early look-say or "thought" methods that teach no phonics. Near them, though closer to the code side, we can place the intrinsic-phonics programs.

This classification is, of course, based on *emphasis*. All code programs, then as now, give practice in reading for meaning, and all meaning programs give practice in the code.

SYNTHESIZING FINDINGS

I noted above that different investigators used performance on different measures of reading success as a basis for concluding that one method was superior to another. Some used accuracy of oral reading; others, speed of reading or degree of comprehension. Some tested spelling; others spoke of "attitude toward reading." Still others tested a variety of outcomes.

I suspected that their conflicting findings were due, at least in part, to this diversity. Thus, to sharpen understanding of what "really occurred," I decided that results for the different measures should be tabulated separately. All aspects of reading do not necessarily improve at the same rate from year to year, and different methods may show different strengths and weaknesses at different times. For example, certain methods

may equip children to perform better on one or two measures in the middle or at the end of grade 1, but when these children are followed up by further testing in later grades, their achievement may be less impressive than that of children who get off to a slower start using a different method.⁴

So for each experimental comparison, any available findings were tabulated in terms of the following eight measures of reading ability, grade by grade:

1. *Word pronunciation*—ability to read (pronounce) words on a list.
2. *Connected oral reading*—ability to read a selection aloud. Oral tests are usually scored only for accuracy, but the score may also be based on a combination of speed and accuracy.
3. *Phonics*—knowledge of letter-sound correspondences.
4. *Spelling*—skill in writing words from dictation. Some of the newer standardized tests measure spelling by having the child choose from among three misspellings of a word and one correct spelling.
5. *Vocabulary*—knowledge of word meaning. Sometimes called “word meaning” or “word reading,” this is usually part of a standardized silent reading test. In grades 1 and 2 such tests usually require the pupil to select the one word out of four that matches a picture. At about grade 3 and above the child reads a “test” word and picks the one word out of four that defines it.

A low score on such a test may mean either that the pupil does not know what the word means or that he cannot identify it (know what it says), or both. Thus, a vocabulary subtest on a standardized silent reading test measures an indeterminate amount of word comprehension, word recognition, and ability to sound out words not seen before. It also contains an element of guessing and test-taking ability. Because there is a time limit, rate of reading also may affect the child's score.

⁴ This is essentially the point Guy Buswell made in 1922 in his *Fundamental Reading Habits: A Study of Their Development*. Buswell's thesis (based on eye-movement records and oral and silent reading tests of first-grade children taught by either a meaning or a code emphasis) was that each person has his own course of growth toward mature reading and that different initial teaching methods produce different patterns of growth. If we assume with Buswell and others that reading is a complex skill, we cannot accept, without further analysis, the assumption that performance on one measure of this skill necessarily represents performance on all other measures—or that one measure is necessarily the most essential at the beginning stage of reading. For example, high speed in the beginning may go with low accuracy. Further, one must consider which measures are most important at certain levels. We all agree that the *end product* of reading should be a high degree of comprehension, adequate rate, and a proper attitude. But do children taught by methods that produce such results at the beginning maintain these skills? This I hoped to learn by breaking down research results by outcomes on the different measures and by successive grades.

6. *Silent reading comprehension*—ability to understand material read. Sometimes called “paragraph reading,” this is usually part of a standardized silent reading test. The child reads selections and answers questions to show that he understands them. Like the vocabulary subtest, the comprehensive subtest measures an indeterminate combination of skills and is affected by rate of reading.

7. *Rate of reading*—how quickly the child reads silently. Accuracy is not usually taken into account, although tests differ considerably on this.

8. *Interest,⁵ fluency, expression*—these were grouped together because only one study (Currier and Duguid, 1916) reported on these measures, although no objective means of judging them was presented.

Look-Say or Phonics?

I found nine studies comparing a look-say with a phonic method—whether systematic or intrinsic. All but one were made before or during the 1930s, when the issue was mainly one of phonics versus no phonics. (After the 1930s, people were asking *how much* and *what kind* of phonics to teach, rather than *whether* to teach it.) Tables 4-1A and 4-1B present the findings of each study by measure tested and grade. In the table, “L-S” indicates that the look-say children scored higher, “SP” indicates that the systematic-phonics children scored higher, and “IP” indicates that the intrinsic-phonics children scored higher on a particular component.

What can we learn from these nine studies? First, the rather shocking fact that no studies tested beyond the second grade; even worse, *most* of the evidence—and it is limited evidence—is for the middle or end of the first grade.

Second, the different methods produced different results depending upon what was tested and when it was tested.

WORD RECOGNITION AND ORAL READING

On word recognition and oral reading, the phonics-trained children were ahead of the look-say children in both the first and second grades.

Valentine (1913) gives us some idea of why a phonic approach produced better results. Dividing test words into “taught” and “untaught”

⁵ Gates (Gates et al., 1926) attempted to measure interest objectively in his comparison of “opportunistic” and “modern systematic” methods. That no one else reported results does not mean that many did not claim superior interest for this or that method.

Table 4-1A Summary of Experiments Comparing Look-Say and Phonic Methods (Advantage in Grade 1)*

Experiment and methods compared	Oral word recognition	Con- nected oral reading	Silent reading		Rate	Interest, fluency, etc.
			Vocab- ulary	Compre- hension		
Gill (1912) (L-S vs. SP)					L-S	
Valentine (1913) (L-S vs. SP)	High IQ SP (taught words) SP (untaught words) Low IQ L-S (taught words) SP (untaught words)					
Currier & Duguid (1916) (L-S vs. SP)	SP			L-S	L-S	L-S
Winch (1925) (L-S vs. SP)	SP	SP			L-S† SP‡	
Sexton & Herron (1928) (L-S vs. SP)			L-S (Jan.) SP	SP		
Tate (1937) (L-S vs. SP)			SP (Apr.)	L-S (Apr.)		
Gates & Russell (1938) (L-S vs. IP vs. SP)			IP (Feb.)	IP (Feb.)		
Tate et al. (1940) (L-S vs. IP)			L-S (March)	L-S (March)		

* Tests given at end of first grade unless otherwise specified. L-S = look-say; SP = systematic phonics; IP = intrinsic phonics.

† Irrespective of accuracy.

‡ Number of words recognized per minute.

(the latter contained the phonic elements taught to the phonics group), he found that children with phonics training generally had an even greater advantage on untaught words than on taught words. This would suggest that a phonic approach has a greater transfer value. A reanalysis of Valentine's results for the low-IQ group indicated that systematic phonics gives children the same advantage on untaught words, although the look-say approach seemed to help this group more with taught words.

COMPREHENSION AND VOCABULARY

On silent reading tests—whether of vocabulary or reading comprehension—results appear to have depended on when the test was given. The earlier the testing, the better the look-say groups performed. But by

Table 4-1B Summary of Experiments Comparing Look-Say and Phonic Methods (Advantage in Grade 2)*

Experiment and methods compared	Oral word recognition	Con- nected oral reading	Silent reading		Rate	Interest, fluency, etc.
			Vocab- ulary	Compre- hension		
Gill (1912) (L-S vs. SP)					L-S	
Valentine (1913) (L-S vs. SP)	High IQ SP (taught words) SP (untaught words) Low IQ L-S (taught words) SP (untaught words)					
Currier & Duguid (1916) (L-S vs. SP)	SP			L-S	L-S	L-S
Sexton & Herron (1928) (L-S vs. SP)				SP (after 5 mos.)		
Mosher & Newhall (1930) (L-S vs. SP)			SP	SP	SP (NS)	
Tate et al. (1940) (L-S vs. IP)			IP	IP		

* Tests given at end of second grade unless otherwise specified. L-S = look-say; SP = systematic phonics; IP = intrinsic phonics; NS = not statistically significant.

the second grade, all except Currier and Duguid (who did not report test data, but only descriptive, qualitative statements) found an advantage for phonics—whether systematic or intrinsic.

I explain this as follows: Standardized silent reading tests are usually timed. They contain high-frequency words, often irregularly spelled. The children following a look-say method probably spent a greater amount of time learning these very words, while the children studying phonics were learning to associate letters with sounds. The look-say children also acquired a “set” for getting meaning; that is, using pictures and context to get the general thought. This helped them score higher on standardized silent reading tests which contain such aids. The phonics-trained children, especially those given more systematic-phonics instruction, acquired a set for accuracy in recognizing words and were preoccupied with letters and sounds (as noted in Currier and Duguid, 1916; Winch, 1925; and Gates and Russell, 1938). This kind of learning set pays off in oral, untimed tests of word recognition, but can lead to lower scores on silent, timed reading tests, especially in grade 1, when the average

phonics-trained child has not yet acquired sufficient skill to "sound out" words that he cannot recognize immediately. This child concentrates on working out the words and has less time and energy left to devote to getting the thoughts. By the second grade, however, when the average phonics-trained child has mastered a sufficient amount of phonics knowledge and skill, he can devote himself more to meaning.

RATE

All studies testing rate of reading (with the exception of Mosher and Newhall, 1930) found that a look-say approach resulted in faster reading in both the first and second grades. One study, however, differentiated between rate measured irrespective of accuracy and number of words recognized accurately in one minute. Winch, testing first graders, found that the look-say group read faster when accuracy was not considered but that the phonics group recognized more words accurately per minute.

Mosher and Newhall found that their systematic-phonics group had a slight (although not significant) advantage in rate at the end of grade 2.

To sum up, our analysis of these look-say versus phonics studies indicates that an initial phonic approach (whether systematic or intrinsic) probably results in lower comprehension and rate at the beginning of grade 1 than look-say, but achieves better results in comprehension by the end of the second grade. A tenable hypothesis is that if given time, phonics is advantageous not only for word recognition but also for comprehension—one of the ultimate goals of reading instruction.

Most of the investigators making these studies—and most summarizers of these studies—did not interpret the findings as I have. The advantage for comprehension that a phonic emphasis produced at the end of the first or second grade was interpreted to mean that phonics should be delayed until then. (This seems a questionable interpretation, since a delayed phonics program was not actually tested. In fact, there is stronger evidence for the hypothesis that *early* training in phonics is required to produce higher scores on comprehension tests at the end of the first or second grade, even though it does not pay off for beginning comprehension. It is probably the cumulative effect that is crucial in producing the later advantage.)

Interestingly, only one study (Currier and Duguid, 1916) reported evidence on interest, fluency, and expression—and only in descriptive and highly subjective terms. The advantage on these, according to the investigators, went to the look-say group.

Systematic or Intrinsic Phonics?

I analyzed twenty-five studies that compared systematic and intrinsic phonics for beginning readers.⁶ These are more recent than the studies discussed above. In fact, more have been reported in the decade between 1955 (when *Why Johnny Can't Read* appeared) and 1965 than in the period 1926 to 1953 (see Tables 4-2A through 4-2F).

Most of the investigators again tested only at the end of the first grade. Fewer made comparisons at the end of the second grade, and a still smaller number tested at the end of the third and fourth grades. Beginning with the fifth grade there are practically no data.

Though the evidence available on this issue is still deplorably limited, these experiments exhibit somewhat more scientific rigor than the older look-say versus phonics experiments. They show an increasing

⁶ Additional studies brought to my attention after the major analysis was completed, as well as the most recent USOE-supported first-grade comparisons, are reviewed briefly later in this chapter.

*Table 4-2A Summary of Experiments Comparing Systematic and Intrinsic Phonics (Advantage in Grade 1)**

Experiment	Oral word recognition	Connected oral reading	Phonics	Spelling	Silent reading	
					Vocabulary	Comprehension
Peyton & Porter (1926)						IP
Gates (1927)	E (Apr. & May)	E (Apr. & May)	E (Apr. & May)		IP (Apr. & May)	IP (Apr. & May)
Garrison & Heard (1931)	SP	IP		SP (bright) E (dull)		SP (bright) E (dull)
Gates & Russell (1938)					IP (Feb.)	IP (Feb.)
Russell (1943)	SP		SP	SP	SP	SP
Wohleber (1953)					SP	SP
Henderson (1955)					SP	SP
Sparks (1956); Sparks & Fay (1957)					SP	SP
Linehan (1958)		SP	SP		SP	SP
Bear (1959 & 1964)			SP		SP	SP
Bloomer (1960)					SP (date not given)	SP (date not given)
Bloomer (1961)					SP	E

* Tests given in May or June unless otherwise specified. SP = systematic phonics; IP = intrinsic phonics; E = equal.

*Table 4-2B Summary of Experiments Comparing Systematic and Intrinsic Phonics (Advantage in Grade 2)**

Experiment	Oral word recognition	Spelling	Silent reading	
			Vocabulary	Comprehension
Garrison & Heard (1931)		SP (bright) E (dull)		E (bright) IP (dull)
Russell (1943)		SP		
Wohleber (1953)			SP	SP
Henderson (1955)		SP	SP	SP
Sparks (1956); Sparks & Fay (1957)			E	SP
Kelly (1958)			SP	SP
Daniels & Diack (1960)	SP			SP
Tensuan & Davis (1963)				IP (NS)
Duncan (1964)		SP	SP	SP

* Tests given in May or June. SP = systematic phonics; IP = intrinsic phonics; NS = not statistically significant; E = equal.

*Table 4-2C Summary of Experiments Comparing Systematic and Intrinsic Phonics (Advantage in Grade 3)**

Experiment	Oral word recognition	Connected oral reading	Spelling	Silent reading		
				Vocabulary	Comprehension	Rate
Garrison & Heard (1931)			SP (bright) SP (dull)		E (bright) SP (dull)	
Agnew (1935 & 1939)	SP	SP		SP	SP (directions) SP (details) E (general significance, outcome, accuracy)	E
Wohleber (1953)				SP	SP	
Henderson (1955)			SP	SP	SP	
Sparks (1956); Sparks & Fay (1957)				E	E	
Gates (1961a)				E or IP†	E or IP†	E or IP†
Morgan & Light (1963)			E	IP	IP (Gates Basic) SP (NS) (Calif.)	
Tensuan & Davis (1963)					IP (NS) in English‡	
Duncan (1964)			SP	SP	SP	

* Tests given in May or June. SP = systematic phonics; IP = intrinsic phonics; NS = not statistically significant; E = equal.

† Several kinds of analyses were performed, two indicating approximately equal results, and one indicating somewhat lower achievement for SP children.

‡ Children were initially taught in Tagalog, the Philippine national language. Reading achievement was tested at the end of grade 2 in Tagalog and at the end of grade 3 in English.

Table 4-2D Summary of Experiments Comparing Systematic and Intrinsic Phonics (Advantage in Grade 4)*

Experiment	Phonics	Silent reading			
		Spelling	Vocabulary	Compre- hension	Rate
McDowell (1953)		SP	IP	IP	IP
Sparks (1956); Sparks & Fay (1957)		E	E	E IP (accuracy)	E
Gates (1961a)			E or IP†	E or IP†	E or IP†
Wollam (1961)		SP	SP	E	SP
Edward (1964)	SP		SP	SP	SP

* Tests given in May or June. SP = systematic phonics; IP = intrinsic phonics; E = equal.
† Several kinds of analyses were performed, two indicating approximately equal results, and one indicating somewhat lower achievement for SP children.

Table 4-2E Summary of Experiments Comparing Systematic and Intrinsic Phonics (Advantage in Grade 5)*

Experiment	Silent reading		
	Vocabulary	Comprehension	Rate
Gates (1961a)	E or IP†	E or IP†	E or IP†

* Tests given in May or June. IP = intrinsic phonics; E = equal.
† Several kinds of analyses were performed, two indicating approximately equal results, and one indicating somewhat lower achievement for SP children.

Table 4-2F Summary of Experiments Comparing Systematic and Intrinsic Phonics (Advantage in Grade 6)*

Experiment	Silent reading			
	Spelling	Vocabulary	Comprehension	Rate
Bear (1959 & 1964)	SP	SP	SP (NS)	SP (NS)

* Tests given in May or June. SP = systematic phonics; NS = not statistically significant.

trend toward the use of standardized silent reading tests. The testing of reading rate almost disappears, with no reports for grades 1 and 2. Some do, however, test for spelling ability. Oral reading—both oral recognition (pronunciation) of isolated words and connected oral reading—receives relatively little attention (probably because by then these measures were already considered minor objectives). These studies do not report at all on interest, fluency, expression, etc.

Tables 4-2A through 4-2F summarize the results by different measures of reading achievement and by grade. As in Tables 4-1A and 4-1B,

I simply indicate the method that had the advantage on a particular test at a particular time, using SP and IP to stand for systematic phonics and intrinsic phonics, respectively; E (equal) indicates no difference in results with the two methods.

WORD RECOGNITION AND ORAL READING

Few investigators tested oral reading, but of those who did, all but one found the systematic-phonics group superior to the intrinsic-phonics group. Gates (1927) differed; he found both groups about equal. This might be because he tested early (in April and May of grade 1), before the systematic-phonics group had learned enough phonics to help them sound out the common, irregularly spelled words on the test.⁷

PHONICS AND SPELLING

The evidence on phonics knowledge is limited to the first grade. All studies (except Gates, 1927)⁸ found that the systematic-phonics groups learned more phonics than the intrinsic-phonics groups.

Tests of spelling ability indicated either equal achievement or superior achievement by systematic-phonics groups—whether spelling was tested in grade 1, 2, 3, 4, or 6.

VOCABULARY AND COMPREHENSION

Most studies reported that children who were exposed to systematic phonics did better on standardized silent vocabulary tests at the end of grades 1, 2, and 3 than children who were learning from an intrinsic-phonics program. Beginning with grade 4, the evidence is more limited and also less clear-cut.

My explanation for this is as follows: Theoretically, the advantage that an early and heavy emphasis on phonics gives a child taking a standardized silent test of vocabulary is probably limited. In the lower grades most of the words on these tests are no doubt well within the meaning vocabularies of most children; thus a child's performance depends mainly on his ability to recognize the word. In the later grades this ability is still important, but differences in comprehension of word

⁷ Recent correspondence with Gates offers another possible explanation. His intrinsic-phonics materials were extremely well programmed, teaching the alphabet along with the words. Thus, both groups probably received similar amounts of decoding practice.

⁸ The reason may be that given in footnote 7.

meanings—i.e., in the child's general knowledge and experience—become more crucial.

The pattern that these studies show for silent reading comprehension is somewhat similar to that for vocabulary. Generally, systematic phonics appears to give more help than intrinsic phonics to the child taking a silent reading comprehension test in grades 1, 2, and 3. The advantage seems, however, to vary with the time of testing. In grade 1, it does not seem as great or as consistent as it does in grade 2. In fact, when children are tested early in grade 1, the intrinsic-phonics group may perform better (Gates, 1927; Gates and Russell, 1938). We can perhaps explain this in terms of learning set (discussed on page 107) and the nature of the standardized reading tests used in the studies.

By the end of grade 2, the systematic-phonics groups seem to have done decidedly better on these tests. By the end of grade 3, although systematic phonics still scores ahead, several studies report either no differences or advantages in certain aspects of comprehension only.

We lack enough evidence to be able to say anything definite about the effect of either systematic or intrinsic phonics on silent reading comprehension in grades 4, 5, and 6, except that the evidence does tend to disprove the long-accepted generalization that systematic phonics produces readers who *do not* read for meaning.

In this measure of reading ability, too, there is probably a limit on any advantage that systematic phonics might impart. In the early grades comprehension depends largely on recognizing words whose meanings are already known. Beginning in about grade 4, a child might be unable to grasp the meaning of, and the ideas conveyed by, all the words he is capable of identifying (i.e., of pronouncing). At this point his breadth of experience, his general intelligence, and other factors probably determine his reading comprehension test scores more than his ability to identify words (see Chapter 5).

RATE

The few studies that tested rate do not indicate any clear-cut trends. A systematic-phonics emphasis seems to have produced slower rates in grades 1 and 2 (see also the studies comparing phonics and look-say, which produced similar results), but not to have affected rate negatively in grades 3 and 4. In fact, two studies provide evidence (admittedly slim) that systematic phonics may contribute to a higher rate by grade 4. At least the existing evidence cannot be used to justify the often-repeated assertion that children who begin with systematic phonics become *slower* readers.

In summary, judging from the studies comparing systematic with intrinsic phonics, we can say that systematic phonics at the very beginning tends to produce generally better reading and spelling achievement than intrinsic phonics, at least through grade 3.

More specifically, the child who begins with systematic phonics achieves early superiority in word recognition. This superior ability may not always show up on standardized silent reading (comprehension and vocabulary) tests in the first grade. But, by the second and third grades, greater facility in recognizing words probably increases his ability to read for meaning, as measured by standardized silent reading tests of vocabulary and comprehension.

As for rate, systematic phonics may produce slower readers in grades 1 and 2 because it develops greater concern for working out the words. However, by the middle grades, rate seems to be about equal to that produced by intrinsic phonics.

Finally, there is probably a limit to the advantage that early facility with the code gives on comprehension tested after grade 4. After this point intelligence, experience, and language maturity probably become more important factors in success than ability to recognize words (see Chapter 5).

skip to p. 133

Additional Studies on Varying Amounts of Phonics

Three studies were called to my attention⁹ after the above analysis was completed.

A 1955 study by the Committee on Research and Guidance of the Department of Public Instruction, Queensland, Australia, compared an experimental group (that I classified as intrinsic phonics) with a control group (systematic phonics). After three years the systematic-phonics group scored higher than the intrinsic-phonics group on tests of comprehension and attack on new words, and somewhat higher on tests of word recognition.

Strange as it may seem, but not unlike some previous researchers, the authors of this study concluded *for* intrinsic phonics, as follows:

The evaluation of this experiment by objective tests does not demonstrate that either method produces clearly superior results.

We are satisfied that the methods used in the experimental [intrinsic phonics] schools have produced intangible benefits that are not readily assessed quantitatively.

⁹ By Theodore Clymer, of the University of Minnesota, and Mrs. Anne Hughes, former research director of the Reading Reform Foundation.

skipped from

P. 114

133

Experimental Evidence

The Novelty Effect

of the three studies concluded that the innovative method, look-say, was better.

When look-say was the accepted method (from 1920 to 1935), in theory at least,²² all the studies concluded for phonics—twice as many for systematic as for intrinsic phonics. But while systematic phonics may have been the “out” method then, albeit only in theory, it had probably not yet been “out” long enough to be innovative again and to bring with it the novelty effects of new methods.

Between 1935 and 1955, when intrinsic phonics was “in,” we find again that the “out” method (systematic phonics) tends to win.

During the period of the debate (1955 to 1965), when intrinsic phonics was still the dominant approach (and systematic phonics old enough to be innovative again), there seems to have been a real buildup of conclusions for systematic phonics.

These trends in conclusions would seem to indicate that some novelty effect is operating. However, the verdicts (which tend on the whole to support a stronger phonic emphasis, irrespective of historical period) reflect the authors' conclusions, some of which ran counter to their own findings or were based on what I consider questionable tests for that particular stage of reading development. Indeed, had they based their conclusions on their own findings, many more investigators would have concluded for systematic phonics.

On the basis of this very limited study, I would say that the novelty

²² A. I. Gates writes: “In your report you seem to state that the decade 1920–30 was the time when the nonphonic or low phonic or the ‘look and say’ approach was in widespread use. I think that this was not the case. Early in my work I made quite an extensive investigation of schools in the New York area by direct observation of teachers in their classrooms, by conferring with teachers from all parts of the country who were members of my college courses, and by analyzing all or nearly all the basal reading books in use at that time. The results were published in ‘Problems in Beginning Reading,’ *Teachers College Record*, March 1925. I found that extensive phonic work was provided and recommended in nearly all the basal reading programs and was in fact being done in practically all schools. This was also shown by a study made by Clifford Woody, *Practices in Teaching First-Grade Reading in the Public Schools of Michigan*, Bureau of Educational Reference and Research Bulletin 58, School of Education, University of Michigan, 1923, who reported that in 1921–2 a quarter of the teachers spent half or more of their time, one half spent 38% to 48%, and three quarters spent 27% to 38% on word study including phonics. Despite the fact that William S. Gray, I, and others, beginning about 1920, were questioning the desirability of devoting so much time to drill on phonics, the phonic method was heavily entrenched in the schools for many years thereafter. Although Nila Smith characterized 1920 as the time when the critical views began to appear, almost everything I wrote and said during the 1920–30 decade was based on the assumption that most teachers were committed to a very heavy program of phonics.” (Personal communication, November, 1965)

effect probably had some influence on conclusions as well as on findings of these studies. However, the more essential factor influencing their results, I believe, was the degree of emphasis on the code afforded by the methods under examination. Even if we choose the most simple-minded way out—i.e., if we take a ballot (indeed, a risky business if no theory or explanation accompanies it)—we find that, except for the period 1900 to 1920, the stronger code-emphasis methods tended to produce the better results.

But the major evidence that a stronger code emphasis at the beginning produces better results than a weaker one (one that emphasizes meaning initially) stems not from “head counting” or “ballot taking” but from my theoretical analysis of the probable course of development of reading skill when children are given more or less instruction in the code system of English writing, discussed more fully above. The studies are used as confirmations or denials of this theoretical position—which, like all theories, will of course need refinement and revision as new evidence comes to the fore.

A recent study by Bliesmer and Yarborough (1965), published after the major portion of this chapter was written, tends to confirm my basic interpretations of the past classroom experiments as well as my judgment that a novelty effect did not have a major influence on their results.

Bliesmer and Yarborough compared ten beginning reading programs that varied in the emphasis given to code learning. Three programs (the Ginn Basic Readers, 1959; American Book, 1963; and the Scott, Foresman Reading Program, 1962) were of the conventional basal-reader type (my intrinsic, moderate phonics or meaning-emphasis classification). Three of the programs had a stronger code emphasis (Houghton Mifflin's *Reading for Meaning Series*, 1963; Economy's *Phonetic Keys to Reading*; and the Lippincott Basic Reading Program, 1963). Also included was a linguistic-type program, Stern's *Structural Reading Series*. Two programs involved an individual approach—one using no set of commercially prepared materials, and one using as supplemental materials the SRA Reading Lab (essentially reading-for-meaning exercises) and word games (code practice).

All the children took the same standardized tests (Stanford Achievement) at the end of grade 1. Those using the programs with heavier code emphasis—whether phonic or linguistic—scored higher than those using the conventional basal-reader programs (more moderate emphasis on the code and greater emphasis on “meaningful reading”) and also higher than those using both Individualized Reading approaches (which also tend to give less systematic practice in code acquisition—see Appendix A). It is important to note that the Stanford test is supposed to measure reading comprehension.

Since the two Individualized Reading programs were also innovative programs at the time of this study, they, as well as the stronger code-emphasis programs, should have been the winners if the novelty effect was stronger than the effect of method. Instead, only the code programs, and not the *new* meaning programs, were superior to the conventional, prevailing meaning programs.

An Addendum on the USOE First-grade Reading Studies

The recently completed twenty-seven USOE first-grade reading studies (*The Reading Teacher*, May, 1966) afford an opportunity to test some of the interpretations of the classroom experiments from 1910 to 1965. Unfortunately, time did not permit my analyzing and synthesizing their findings with those completed before 1965. Nor were all these studies available at the time of this writing. Therefore, I rely essentially on the progress report of Guy Bond (1966), the coordinator of all twenty-seven studies. In Bond's words:

The major goal of the First-grade Reading Study is to explore the effects upon early reading growth of various approaches to reading under conditions that would make it possible to compare findings among a group of independent studies. As the result of a widely publicized invitation by the U.S. Office of Education, 76 proposals were received. Each research proposal submitted was reviewed by the Research Advisory Council which selected 27 for support by the Cooperative Research Branch of the United States Office of Education. The projects were selected on the basis of their individual merit as self-contained studies having unique characteristics and at the same time being directly related to the problems and differing points of view in regard to initial reading instruction. They were also selected so that in total scope a more massive body of information about various approaches to reading instruction could be obtained for further combined analysis than had ever before been possible in the field of reading. (p. 2)

Although these were twenty-seven independent studies, extending over almost all the United States, they were coordinated in terms of research design, measuring instruments, and information gathered, so that comparisons among the studies were possible "in ways that have not previously existed." (p. 2)

The major group of studies investigated the effectiveness of a variety of approaches

... including those employing [conventional] basal readers, phonetic [phonic] emphasis, linguistic materials, language-experience approaches, the Initial Teaching Alphabet, and diacritical markings. (p. 2)

The Coordinating Center at the University of Minnesota, under Bond's direction, treated the data from all the individual studies as one large study.

Although the statistical analyses had not yet been completed when Bond wrote his progress report, he presented "certain specific generalizations that can be made at this time on the basis of the analyses of the combined data thus far completed." (p. 8)

His first generalization confirms one that I made from the past studies:

We have found no one approach²³ so distinctly better in all situations and respects than the others that it should be considered the one best method nor to be used exclusively. (p. 8)

Bond generalizes further:

There are, however, many indications that no matter what the underlying method is, word-study skills need to be emphasized and taught systematically. This is best shown by the superiority of the approaches which augmented the basal reader with a phonetic [phonic] emphasis as compared with basal readers as usually taught. (p. 9, emphasis mine)

The above seems to confirm my interpretation of the past research—i.e., programs that facilitate code acquisition tend to produce generally better results at the end of grade 1.

The USOE studies also found that some code-emphasis programs may not show their advantage on comprehension tests at the end of grade 1, although they do so on word-recognition tests:

The combined data analysis tends to show that a linguistic approach develops word-recognition skills effectively, but demonstrates no superiority over other systems in developing comprehension abilities. (p. 9)

This finding is reminiscent of the earlier ones already analyzed in this chapter. Now it remains to be seen whether these "linguistic-trained" children will catch up in reading comprehension and, if they do, when?

It will also be important to see whether the second-grade studies, scheduled to be completed in 1967, will necessitate a modification of the generalizations from the earlier studies presented in the following summary.

A Summary, Particularly for Researchers on Method

We have viewed the experimental studies analyzed in this chapter as sources of evidence on the basic question: Does an initial code emphasis produce better results than an initial emphasis on reading for meaning?

²³ Bond refers here to the specific programs rather than to my basic categories of meaning emphasis and code emphasis.

I cannot stress sufficiently that this dichotomy is only one of *emphasis*. All code-emphasis programs give some practice in reading for meaning during the initial stages. Many put great stress on it. Reading for meaning really cannot be avoided unless, of course, only nonsense syllables are used, and no code-emphasis program has ever gone to that extreme.

Similarly, the meaning-emphasis programs give some practice in code learning. Reading could not take place without some attention to the code. But compared with code-emphasis programs, the meaning-emphasis programs provide less practice, and they give it later.

When we view the analyses given above in terms of this code-meaning dichotomy, the following generalizations—which I present as hypotheses to be tested further—seem tenable:

1. A code emphasis tends to produce better overall reading achievement by the beginning of fourth grade than a meaning emphasis.

2. Growth in reading skills seems to take different courses under the two emphases.

- a. Under a code emphasis, the child shows, from the very beginning, greater accuracy in word recognition and oral reading; this may or may not give him an immediate advantage on reading-for-meaning tests (standardized silent reading tests of vocabulary and comprehension). However, by the end of the first or sometime during the second grade, the early advantage in word recognition produces better vocabulary and comprehension scores on silent reading tests. These advantages persist through about the third grade.

With a code emphasis, the child seems to read more slowly at the very beginning because of the greater stress on accuracy. However, by the third (or fourth) grade, when he is more fluent, his rate is equal to (and may ultimately exceed) that produced by a meaning emphasis.

- b. Under a meaning emphasis, the child has an early advantage (in the middle of grade 1) on reading-for-meaning tests (standardized silent reading tests of vocabulary and comprehension). However, he has an early disadvantage in accuracy or oral word recognition (pronunciation) and connected oral reading tests (when rate is not included in the score), which ultimately dissipates the early advantage on the standardized silent reading tests. At about the end of the first grade (or the beginning of the second grade), and continuing through about the third grade, meaning-emphasis programs tend to **affect** comprehension and vocabulary test scores adversely, mainly because the child does less well in word recognition.

In the beginning, the child reads faster under a meaning emphasis, but he may lose this advantage by about the third or fourth grade.

3. There is more than one way to facilitate learning of the code; systematic-phonics programs that rely on direct teaching of letter-sound relationships are as successful as, or perhaps more successful than, programs that rely on "discovery"—the so-called linguistic approaches that *do not* teach letter-sound correspondences directly.

4. Modified alphabet schemes and linguistic approaches that control words on spelling patterns to permit discovery of letter-sound correspondences tend to help the child master the code. They are probably superior to approaches that expose the child to high-frequency, irregularly spelled words with "late and little phonics" (the phonics programs of the conventional basal readers). However, even the modified alphabet schemes and the linguistic approaches appear to benefit from direct teaching of sound-letter correspondences.

5. With regard to individual differences:

a. Children of below-average and average intelligence and children of lower socioeconomic background probably learn better in the end with a code emphasis than with a meaning emphasis, although this advantage does not show immediately.

b. Children of high mental ability and children of middle and high socioeconomic background appear to gain an immediate advantage from a code emphasis. However, because they are bright they are generally better able to discover sound-letter relationships for themselves. Thus, the differences between results from a meaning or a code emphasis are probably not ultimately as great for them as for average and slow learners and for children of lower socioeconomic background.

The experimental comparisons do not provide enough evidence to generate tenable hypotheses about differential growth patterns at the fourth grade and beyond. However, I would like to present the following set of second-order hypotheses:

1. Whether an initial code emphasis keeps its advantage in the middle and upper elementary grades, and later, depends on how reading is taught in these grades: how much the reading program stresses language and vocabulary growth and provides sufficiently challenging reading materials. If the reading programs are not challenging enough in these respects, the early advantages will probably be dissipated.

2. Generally, aspects of reading comprehension such as "reading to predict outcomes," "making inferences," "reading for appreciation," and the like may not show substantial differences in later years when initial mean-

ing and code programs are compared, since the reader's intelligence and general knowledge put a limit on performance in these areas. However, a code emphasis should still maintain its advantage, even in later years, in those aspects of literacy which depend less on language, intelligence, and experience and more on "reading skill": accuracy in recognizing "unknown" words, accuracy and rate of connected oral reading, rate of silent reading, and some kinds of reading comprehension—e.g., reading for details and following directions. A code emphasis will tend to maintain its early advantage in spelling.

It should be made clear that the above hypotheses concern groups, not individuals. Obviously, every method produces ranges of attainment, and every method has its failures. And it may very well be that certain individuals find one or another method particularly suitable—or impossible.