

Defending Whole Language: The Limits of Phonics Instruction and the Efficacy of Whole Language Instruction

Stephen Krashen

Reading Improvement 39 (1): 32-42, 2002

The Reading Wars show no signs of stopping. There appear to be two factions: Those who support the Skill-Building hypothesis and those who support the Comprehension Hypothesis. The former claim that literacy is developed from the bottom up; the child learns to read by first learning to read outloud, by learning sound-spelling correspondences. This is done through explicit instruction, practice, and correction. This knowledge is first applied to words. Ultimately, the child uses this ability to read larger texts, as the knowledge of sound-spelling correspondences becomes automatic. According to this view, real reading of interesting texts is helpful only to the extent that it helps children "practice their skills."

The Comprehension Hypothesis claims that we learn to read by understanding messages on the page; we "learn to read by reading" (Goodman, 1982; Smith, 1994). Reading pedagogy, according to the Comprehension Hypothesis, focuses on providing students with interesting, comprehensible texts, and the job of the teacher is to help children read these texts, that is, help make them comprehensible. The direct teaching of "skills" is helpful only when it makes texts more comprehensible.

The Comprehension Hypothesis also claims that reading is the source of much of our vocabulary knowledge, writing style, advanced grammatical competence, and spelling. It is also the source of most of our knowledge of phonics.

Whole Language

The term "whole language" does not refer only to providing interesting comprehensible texts and helping children understand less comprehensible texts. It involves instilling a love of literature, problem-solving and critical thinking, collaboration, authenticity, personalized learning, and much more (Goodman, Bird, and Goodman, 1991). In terms of the process of literacy development, however, the Comprehension Hypothesis is a central part of whole language.

In this paper I examine some recent research dealing with two fundamental points of contention between the two sides of this debate.

1. The complexity issue: Whole language advocates claim that the rules of phonics are complex and have numerous exceptions. For this reason many are unteachable (Smith, 1994). Skill-building advocates claim that this is not the case. Shanahan (2001), for example, defends giving phonics instruction a major role in reading instruction because "more than 90 percent of English words are phonetically regular" (p. 70). He does not, however, cite research supporting this claim.

2. The method comparison issue: Skill-Building advocates claim that those in phonics-based classes outperform those in whole language classes (National Reading Panel, 2000). Whole language advocates argue that when whole language is defined correctly, when it includes real reading, students in these classes do better on test of reading comprehension, with no difference on skills tests (Krashen, 1999).

The Complexity Argument: Johnson (2001)

Clymer (1963, 1966) investigated 45 phonic generalizations of words in four basal series and concluded that many did not work very well. This result has been a central part of the argument against over-teaching phonics. Here are two well-known examples: The rule "when two vowels go walking the first does the talking" (when two vowels appear side by side, the long sound of the first is heard and the second is silent, as in "bead") worked in only 45% of the cases Clymer examined, and the final e rule (first vowel is long, final e is silent, as in "cake") worked in only 63% of the cases.

Johnson (2001) re-examined Clymer's conclusions. On reading the title of her paper ("The utility of phonics generalizations: Let's take another look at Clymer's conclusions") and the short summary under the title ("English orthography is not easily reduced to a few rules, but there are some general recommendations for teaching about vowels that can be helpful."), one gets the impression that Clymer's results will be contested, and that a new case for direct phonics instruction will emerge from this article. Johnson promises to review what she considers neglected studies that followed Clymer, and promises to present a new analysis.

The neglected studies, however, replicate Clymer's results, with only a few alterations, a conclusion Adams (1990) also arrived at in her discussion of Clymer's work. And Johnson's new analysis confirms that extensive phonics teaching is a hopeless endeavor. She provides, in fact, dramatic evidence that English phonics is extremely complex, which was, in fact, Clymer's point.

Here is an example. As noted above, Clymer concluded that the "two vowels

walking" rule applied only 45% of the time. In her re-analysis, Johnson concludes that this rule works well for five two-vowel combinations: ay (96.4%), oa (95%), ee (95.9%), ai (75%) and ey (77%). Of course, one can easily dispute that 75 or 77% accuracy is enough to justify this conclusion, but more serious is the fact that the situation is a disaster for the 14 other two-vowel combinations Johnson presents in her table 2. While four other two-vowel pairs are regular, none of these meet the criteria of the "walking" rule. Four additional pairs have two possible pronunciations and four more pairs have three possible pronunciations. Finally, another two pairs with two possible pronunciations were considered "very rare." (One of them appears in "fruit," "suit," and "build" and the other in "Asia," "piano," and "official," hardly arcane words.) Very few of these alternative pronunciations follow the "walking rule."

As noted earlier, Clymer concluded that the final e rule worked only 63% of the time. Johnson concludes that for some combinations, it does a bit better (a-o, as in "cake", 77.7%, i-e as in "five," 74.2%, u-e, as in "rule," 76.9%) but it is less efficient for others (for o-e, as in "stove," 58.4%, e-e, as in "these," 16.6%). She concludes that this rule is "surprisingly reliable when restated" (p. 139) and that teachers can teach this rule "with confidence" (p. 138) as long as they encourage a "flexible strategy." This does not seem to be a real step forward: we are trading a simple rule that works 63% of the time for a far more complex rule that is only slightly more efficient.

Johnson's work, in other words, is a strong confirmation of Clymer's: Simple rules don't apply to a large percentage of words. Yes, one can come up with rules that cover more words, but they also have numerous exceptions and are very complex.

Johnson has no clear program for teaching sound-spelling correspondences, other than the suggestion that we also consider rules for higher order units, such as rimes and "vowel patterns," also, as she points out, numerous and complex.

Buried deep in Johnson's article is the suggestion that some children can acquire phonics generalizations by reading. As noted earlier, Smith (e.g. 1994) has hypothesized that most of our knowledge of phonics is the result of reading and not the cause. Johnson's view differs somewhat from Smith's in that she claims that some children can indeed acquire sound-spelling correspondences by reading, while others "need systematic instruction" (p. 141). No evidence is provided for this extremely important claim, a claim that runs counter to current official state and federal government policy that all children must have systematic, intensive phonics instruction.

To support such a claim, one would have to show that there are substantial numbers

of children who have learned to read without extensive phonics training (this is easy to find), and also substantial numbers of children who cannot "learn to read by reading," who require extensive phonics instruction. The existence of this second group has never been demonstrated: To do so, one must find large numbers of children who have been read to, who have substantial exposure to comprehensible and interesting texts, and who nevertheless fail to learn to read.

One sees, of course, some children who learn to read less quickly than others do, but this is a statistical necessity in any phenomena that exhibits any degree of variability. No matter what, we will always have children who fall in the lowest 25% in rate of learning to read. What we very rarely see are those who never learn to read despite the availability of comprehensible and interesting print.

Which phonics generalizations are useful? Which ones really help children understand text? I suggest we ask the real experts: Teachers who have helped children learn to read for many years. A consensus of experienced practitioners will tell us if it is worthwhile to tell children that the a-e combination is pronounced with the long vowel and the final e silent (except when the final syllable is unaccented - then the vowel is pronounced with a short-i sound, as in "palace," or the combination is "are," with words such as "have" and "dance" as exceptions). How many of us who easily and fluently read words with the a-e combination were ever aware of this rule?¹

The Method Comparison Argument

As noted earlier, each side has claimed victory in method comparison studies. Skill-building advocates claim that children in skills-based classes learn to read better, while whole language advocates claim that whole language is superior, as long as it is defined correctly. I discuss here a recent contribution to this debate.

Jeynes and Littell (2000) reviewed 14 studies and concluded that overall, low SES children do not benefit from whole language instruction, but "there may be some advantages to the whole language approach in its purist form" (p. 21). Of the 14 studies, only four were listed as published in journals or books. Of the ten studies that were listed as unpublished, two, it turned out, were in fact published in the Reading Research Quarterly. I was able to locate five others through ERIC, and one other that I believe to be identical to a study on Jeynes and Littell's list. My interpretation of these studies is quite different than Jeynes and Littell's interpretations in most cases.

As usual, the definition of whole language is at issue. Jeynes and Littell classified two studies as "pure" whole language. They satisfied the following criteria: (1) no adapted texts; (2) no whole class, teacher sponsored assignments, (3) "integrated language experiences as opposed to direct instruction in isolated skill sequences" (p. 23). The two studies in this category (Manning, Manning and Long, 1989; Morrow, 1992) showed the strongest advantage for whole language. Less "pure" versions of whole language resulted in weaker and negative results.

Jeynes and Littell also classified several "language experience" treatments as whole language, considering language experience to be a "precursor to whole language," p. 27. The core of language experience consists of students dictating stories to teachers; these stories are transcribed by the teacher and used as reading material.

I object to Jeynes and Littell's definition of whole language. (1) In my view, the issue is not whether texts are adapted or modified but whether they are interesting and comprehensible. (2) There are some instances when a whole class teacher sponsored assignment or activity is appropriate in a whole language class. (3) Some phonics knowledge can help make texts comprehensible (Smith, 1994). While most whole language proponents prefer to teach phonics in context, I know of no reason why integrated versus isolated teaching of phonics should be part of the core definition of whole language. The real issue is whether texts are comprehensible.

A similar analysis, limited to published studies, appeared in Krashen (1999), examining studies that compared "whole language" and "skills." This analysis focused only on one characteristic: The amount of real reading for meaning done by the children. As noted earlier, comprehensible and interesting reading is not the only characteristic of whole language, but it is at the core of whole language. The conclusion was that children in classes with more real reading tended to do better on tests of reading comprehension, read more, liked reading more, and did just as well as "skills" students on skills tests (reading nonsense words). This analysis included some of the published studies that Jeynes and Littell included.

I present here comments on those studies in Jeynes and Littell I was able to obtain. As in Krashen (1999), I focus primarily on performance on tests of reading comprehension. Reading comprehension is, after all, the goal of reading instruction.

The impact of whole language/language experience was measured by Jeynes and Littell using effect sizes. Effect sizes are usually calculated by subtracting the mean of the comparison group from the mean of the experimental group, then dividing the result by the pooled standard deviation. They are weighted for sample size (see e.g. Wolf, 1986). Effect sizes can also be computed from other statistics, such as F, t, and r. Following Jeynes and Littell, in this paper, positive effect sizes indicate an

advantage for whole language or language experience, and negative effect sizes indicate that comparisons did better.

1. Jeynes and Littell included an unpublished dissertation by J.R. Hoffman, which was not available to me. I did, however, find a study by Carline and Hoffman (1976) with a nearly identical title that was obviously the same study. Although Jeynes and Littell calculated an effect size of $-.23$ favoring the conventional reading approach over language experience, Carline and Hoffman (1976) concluded that "teachers who use the language experience approach to reading more often ... showed an increase of 2.9 raw score points more on English reading standardized test scores than those teachers who use it less frequently ($p < .05$)" (p. 43). Carline and Hoffman reported a correlation of $.32$ between the amount of language experience used and student gains in English reading, equivalent to an effect size (d) of $.64$.

2. McCanne (1966) compared the impact of a basal reader approach, an audio-lingual approach, and language experience on Spanish speaking children learning English as a second language. McCanne noted that the language experience approach was not designed for students who are acquiring English as a second language (p. 75). Some modifications in language experience were made in this study; nevertheless, the use of language experience clearly required considerable speaking competence before substantial listening and reading took place, a procedure that violates what is known about language acquisition (Krashen, 1985).

McCanne's results depend on the kind of statistical procedure used. When reading test scores were adjusted for factors such as listening ability in English, measures of cognitive development, SES, measures of teacher competence, and pupil attendance, the basal method was better ($d = -.65$, according to my calculations; based on standard scores presented in McCanne's table 12 (hypothesis six). When raw means were used, the language experience students were better ($d = .36$). It is not clear to me how Jeynes and Littell arrived at their figure of $-.34$.

3. Lamb (1972) earned a substantial $-.75$ effect size in Jeynes and Littell, in favor of the comparison group over whole language. Lamb noted however that all five teachers who used the language experience approach had not used it before and participated in monthly training sessions. The entire duration of the study was only four months.

Contrary to Jeynes and Littell's findings, the results were not clearly in favor of the comparisons. Lamb did several different analyses: Analyses of covariance, controlling for IQ scores, teacher experience, and teacher background showed language experience to be superior (for boys, $d = .47$; for girls, $d = .41$) but a one

way ANOVA with no control for potential confounding variables found the basal method to be better. I was able to calculate effect sizes based on unadjusted (ANOVA) scores for girls ($d = .29$) but not for boys. From inspection of the F ratio, it was clear that the effect was less for boys. It was not clear to me how Jeynes and Littell calculated an effect size of $-.75$ in favor of the basal method. Apparently, Jeynes and Littell used only the simple ANOVA results. Note that in McCanne (1976), language experience did better on with raw scores (ANOVA) and worse on adjusted scores (ANCOVA), the opposite pattern, but in that study, Jeynes and Littell appear to have utilized the ANCOVA result. In my summary below I use the adjusted results.

4. Jeynes and Littell included Ewoldt (1976), a comparison of Follow-Through and Non Follow-Through classes on a story retelling task and reported an effect size of $.05$ favoring language experience. The Follow-Through Model advocates the language experience approach. A serious problem, however, was that only eight of the Follow Through subjects (out of 36) actually came from classes in which the basic program was language experience (see table 3, page 9). Nine others came from classes that used some language experience as supplemental activities. No information was available about the classes eight of the students participated in, and eight others came from classes that included no language experience. Clearly, this study should not have been included.

5. In Usova and Usova (1993), the number of students in the whole language class was small ($n = 8$) and the treatment was a combination of a wide variety of activities, emphasizing the combination of art activities with language arts, and also including reading, hearing stories, writing and direct instruction. We are provided with no information whatsoever on the activities of the comparison group, and we have no idea if experimentals and controls differed with respect to the amount of reading done. Neither experimentals nor controls made much progress over the academic year in reading comprehension, experimentals gaining less than one point on a standardized reading test, and controls actually getting slightly worse.

6. Stallings (1975) was awarded a $-.79$ effect size for comparisons over whole language, but it is difficult to determine how Jeynes and Littell arrived at this figure. This study was not a comparison of methods, but sought to determine predictors of achievement in first and third grade classrooms based on one day of observation. Amount of reading done was not one of the predictors considered.

7. Jeynes and Littell included Harris, Serwer and Gold (1966). I obtained Harris and Serwer (1966a), which appears to cover exactly the same data. Harris and Serwer (1996b) is a shorter version published in the Reading Research Quarterly.

Jeynes and Littell reported an effect size of $-.51$ in favor of comparisons ("skills-centered") over language experience. From Harris and Serwer's table 26, I computed an effect size of $-.18$ in favor of comparisons, based on the Stanford Paragraph Meaning test. What is crucial, however, is that Harris and Serwer report that children in the basal group actually spent more time in reading activities than did the children in language experience. Children in the skills classes spent 56% of the instructional time in reading activities, while children in language experience classes spent only 39.5% of the time on reading activities. It is thus quite likely that the comparison children did more real reading. Thus, both the sign and size of the effect size should be changed for this study. Moreover, Harris and Serwer (1966b) reported positive correlations between the amount of time spent in reading activities and scores on the reading tests ($r = .47$ for paragraph reading).

8. Dahl and Freppon (1994) (also available as Dahl and Freppon, 1995) earned an effect size of $.67$ in favor of whole language. This figure represents a combination of six different measures of literacy development, including tests of concepts about print, the alphabetic principle, story retelling, and concepts about writing. None were measures of reading comprehension.

A closer look at the results showed that the only significant difference between the whole language and skills children was on a task in which the child pretended to read a picture book without words. The story was rated for the presence of aspects of the written narrative register. I calculated an effect size of 1.79 in favor of whole language for this task.

9. Jeynes and Littell calculated an effect size of $.50$ for Morrow, O'Connor, and Smith (1990) in favor of whole language. Based only the standardized test used in the study, I calculated an effect size of $d = -.18$ in favor of the skills group over whole language. The test, however, did not contain a measure of reading comprehension. Literature-based students showed more interest in reading: they could name more authors, took more books home to read, reported more reading at home, and named significantly more kinds of reading material. They were also significantly better on an "attempted reading" test, showing more reading-like behaviors.

10. For Morrow (1992), Jeynes and Littell calculated an effect size of 1.24 . For the two reading comprehension tests included in the study, I calculated effect sizes of 1.84 and $.62$, with a mean of 1.23 , nearly identical to Jeynes and Littell's results. It is clear that the whole language ("literature-based") group read more. Children in the literature-based group spent about 3.5 hours per week with basals and about four hours with literature. They were read to daily, engaged in at least three "literacy activities" per week (e.g. retelling and rewriting stories, book sharing, keeping track of what they read), and had at least three sessions per week in a comfortable

"literacy center" for 30 minutes at a time, during which time they read, wrote and performed stories. Comparison students were read to no more than twice a week and focused nearly entirely on the basal and workbook. Free reading was allowed only when children had finished their basal seatwork.

11. Manning, Manning and Long (1989) lasted three years (k to grade 2), and researchers made biweekly visits to classes to "verify the continuity of the two different literacy programs" (p. 5). Both first and second grade whole language students were significantly better in reading comprehension ($d = 1.97$ in both cases, effect sizes were calculated from p values resulting from Mann-Whitney U's). Nine out of 11 whole language children could name a favorite author at the end of grade 2. None of the skills taught children could.

The small sample size ($n = 11$ in each group) is an obvious weakness of this study, but the care for fidelity of treatment and long duration are obvious strengths.

Method Comparisons: Conclusion

Several studies should not have been included in the meta-analysis. In some, it was not at all clear that there was a genuine comparison of whole language/language experience and basal/skill-oriented methods. These include Stallings (1975), which was not a method comparison at all, Ewoldt (1976), in which many "language experience" subjects did not actually have language experience instruction, Usova and Usova (1993), which included a wide variety of activities under "whole language" and had a very small sample size, and McCanne (1966), which used language experience for second language acquirers, which may have been inappropriate because of premature production demands. Applying my criteria of using only tests of reading comprehension eliminates Morrow, O'Connor and Smith (1990) and Dahl and Freppon (1995) as well.

Of the five studies remaining, in three cases I reached conclusions opposite to those of Jeynes and Littell: I have argued that results were inaccurately reported in Carline and Hoffman (1976) and Lamb (1972) and that the direction of the effect size should be reversed in Harris and Serwer (Harris, Serwer and Gold, 1966). If my interpretations are correct, effect sizes for the still-eligible studies should be changed: for Hoffman, from $-.23$ to $+.68$; for Lamb, from $-.75$ to $+.44$; for Harris and Serwer, from $-.51$ to $.18$, for Manning et. al., from 1.21 to 1.97 . The effect size for Morrow (1992) remains the same.

Only two studies in this set provide clear evidence that one group did more reading

than the other, and in both cases those who read more did better on tests of reading comprehension; Morrow (1992), with an effect size of 1.23, and Harris and Serwer (1966a, 1966b), with an effect size of .18. For other studies in which groups are labeled whole language or language experience, with no clear data on amount read, effect sizes still favor whole language: Carline and Hoffman (1976), with an effect size of .68, Lamb (1972), an effect size of .44, and Manning et. al., (1989), with an effect size of 1.97. The average effect size for all five of these studies is +.90 favoring whole language/language experience. Excluding Manning et. al., the study with the largest effect size, the average effect size is +.63. Even if we include McCanne (1966) and exclude Manning et. al., the average is +.38 in favor of whole language/language experience.

Note that this conclusion is not dependent on my policy of limiting measures to tests of reading comprehension. Allowing Morrow, O'Connor and Smith (1990) ($d = -.18$) and Dahl and Freppon (1995) ($d = .67$) into the re-analysis does not change the final result very much.

Jeynes and Littell's conclusion was that although "pure" whole language students did well, basal/skill groups were in general a winner over whole language, with a mean effect size of -.65 in favor of the comparison groups. My conclusion is nearly exactly the opposite.

In my view, neither of our results should be taken as definitive. No study considered the amount of real reading done to be a central variable, and only two studies attempted to determine the amount of reading students did. In light of Harris and Serwer's finding (see also Evans and Carr, 1985) that children in skills classes actually spent more time reading than children in language experience, one must be cautious in concluding that children in any language experience or whole language class actually read more than those in traditional classes.

What is clear, however, is that Jeynes and Littell's interpretation of the research is not the only possible one.

Overall Discussion and Conclusion

A detailed reading of Johnson (2001) reveals that Clymer's original conclusions stand: the rules of phonics, at least those reviewed by Johnson, remain enormously complex. A close analysis of the actual studies reviewed by Jeynes and Littell (2000) shows that when tests of reading comprehension are considered, when real reading is considered as the core element of whole language, and when details of

studies are examined closely, whole language does very well in method comparison studies. Although the authors conclude otherwise, these studies actually provide evidence for the limits of phonics instruction and the efficacy of whole language.

Note

1. Johnson cites Becoming a Nation of Readers (Anderson, Hiebert, Scott, and Wilkinson, 1985) as holding the position that phonics can be reduced to a few simple rules that can easily be taught by the end of grade two. Not quite. Consider this excerpt:

"... phonics instruction should aim to teach only the most important and regular of letter-to-sound relationships ... once the basic relationships have been taught, the best way to get children to refine and extend their knowledge of letter-sound correspondences is through repeated opportunities to read. If this position is correct, then much phonics instruction is overly subtle and probably unproductive" (Becoming a Nation of Readers, p. 38).

Weaver (1994) and Goodman (1993) have pointed out that Becoming a Nation of Readers sees direct phonics instruction as playing a limited role. In fact, the position presented in the excerpt presented above is not very different from that proposed by Frank Smith (1994): A few straight-forward rules of phonics can be taught directly and can be useful in making texts more comprehensible, but most phonics is the result of reading, not the cause.

References

Adams, M. (1990). *Beginning to Read*. Cambridge, MA: MIT Press.

Anderson, R., Hiebert, E., Scott, J., & Wilkinson, I. (1985). *Becoming a Nation of Readers: The Report of the Commission on Reading*. Washington, D.C.: National Institute of Education.

Carline, D. and Hoffman, J. (1976). *Comparison of Language Experience Approach to Reading with a Conventional Reading Approach in Eight Summer Migrant Schools*. ED 135557.

Clymer, T. (1963/1966). The utility of phonics generalizations in the primary grades. *The Reading Teacher*, 16/50, 252-258/182-185.

Dahl, K. and Freppon, P. (1995). A comparison of inner-city children's interpretations of reading and writing instruction in the early grades in skills-

based and whole language classrooms. *Reading Research Quarterly*, 30(1), 50-74. Published also as Dahl and Freppon, 1994, ED 370 075.

Evans, M, and Carr, T. 1985. Cognitive abilities, conditions of learning, and the early development of reading skill. *Reading Research Quarterly*, 20 (3), 327- 350.

Ewoldt, C. (1976). Miscue Analysis of the Reading of Third Grade Follow Through and Non Follow Through Children in Wichita, Kansas. ED 136 219.

Goodman, K. (1982). *Language, Literacy and Learning*. London: Routledge Kagan Paul.

Goodman, K. (1993). *Phonics Phacts*. Portsmouth, NH: Heinemann Publishing Company.

Goodman, K. Bird, L., and Goodman, Y. (1991). *The Whole Language Catalog*. Santa Rosa, CA: American School Publishers.

Harris, A. and Serwer, B. (1966a). *Comparison of Reading Approaches in First-Grade Teaching with Disadvantaged Children (The Craft Project)*. ED 010 037

Harris, A. and Serwer, B. (1966b). The CRAFT project: instructional time in reading research. *Reading Research Quarterly*, 2, 27-57.

Jeynes, W. and Littell, S. (2000). A meta-analysis of studies examining the effect of whole language instruction on the literacy of low-SES students. *Elementary School Journal*, 101 (1), 21-33.

Johnson, F. (2001). The utility of phonics generalizations: Let's take another look at Clymer's conclusions. *The Reading Teacher*, 55, 132-143.

Krashen, S. (1985). *The Input Hypothesis*. Beverly Hills: Laredo Publishing Company.

Krashen, S. (1999). *Three Arguments Against Whole Language and Why They are Wrong*. Portsmouth, NH: Heinemann.

Lamb, P. (1972). *The Language Experience Approach for Teaching Beginning Reading to Culturally Disadvantaged Pupils*. ED 059 858.

Manning, M., Manning, G., and Long, R. (1989). Effects of a whole language and a skills-oriented program on the literacy development of inner city primary children. ED 324642.

McCanne, R. (1966). A Study of Approaches to First-Grade English reading Instruction for Children From Spanish-Speaking Homes. ED 010 380.

Morrow, L. (1992). The impact of a literature-based program on literacy achievement, use of literature, and attitudes of children from minority backgrounds. *Reading Research Quarterly*, 27(3), 250-275.

Morrow, L., O'Connor, E. and Smith, J. (1990). Effects of a story reading program on the literacy development of at-risk kindergarten children. *Journal of Reading Behavior*, 22, 255-273.

National Reading Panel. (2000). *Teaching Children to Read: An Evidence-Based Assessment of the Scientific Research Literature on Reading and Its Implications for Reading Instruction*. Washington: National Institute of Child Health and Human Development.

Shanahan, T. (2001). Response to Elaine Garan. *Language Arts*, 79 (1), 70-71.

Smith, F. (1994). *Understanding Reading*. Fifth Edition. Hillsdale, NJ: Erlbaum.

Stallings, J. (1975). Implementation and child effects of teaching practices in follow through classrooms. *Monographs of the Society for Research in Child Development*, 40, 7-8 (Serial No. 163).

Usova, C. and Usova, G. (1993). Integrating art and language arts for first grade at-risk children. *Reading Improvement*, 30, 117-121.

Weaver, C. (1994). *Reading Process and Practice*. Portsmouth, N.H.: Heinemann Publishing Company

Wolf, F. (1986). *Meta-Analysis*. Beverly Hills: Sage Publications.